General Internal Medicine
Boston University School of Medicine
2010 Publications – A-L

1. **Alford DP**, Krebs EE, Chen IA, Nicolaidis C, Bair MJ, **Liebschutz J**. Update in pain medicine. J Gen Intern Med. 2010 July;25(11):1222-6. doi: 10.1007/s11606-010-1452-4. PMID: 20632120.

2. **Amaral-Sabadini MB**, **Saitz R**, Souza-Formigoni MLO. Do attitudes about unhealthy alcohol and other drug (AOD) use impact primary care professionals' readiness to implement AOD-related preventive care? Drug Alcohol Rev. 2010;29(6):655-61. PMCID: PMC2965619.

3. Ayotte, BJ**, Kressin, NR.** Race differences in cardiac catheterization: The role of social contextual variables. J Gen Intern Med. 2010;25(8):814-18. PMID: 20383600.

4. **Battaglia TA**, **Santana MC**, **Bak S**, Gokhale M, Lash TL, **Ash AS**, **Kalish R**, Tringale S, Taylor JO, **Freund KM**. Predictors of timely follow-up after abnormal cancer screening among women seeking care at urban community health centers. Cancer. 2010 Feb;116(4):913-21. PMID: 20052731.

5. **Berlowitz DR**, Franklin SS. The clock is ticking: the case for achieving more rapid control of hypertension. J Clin Hypertens. 2010;12(5):323-27. PMID: 20546372.

6. **Berlowitz DR**, Hickey EC, Saliba D. Can administrative data identify active diagnoses for long-term care resident assessment? J Rehab Res Dev. 2010;47(8):719-24. PMID: 21110247.

7. **Berlowitz DR**, Stineman M. Risk adjustment in rehabilitation quality improvement. Topics in Stroke Rehabilitation. 2010;17:252-61. PMID: 20826413.

8. Bertholet N, **Cheng DM**, Palfai TP, **Saitz R**. Factors associated with favorable drinking outcome 12 months after hospitalization in a prospective cohort study of inpatients with unhealthy alcohol use. J Gen Intern Med. 2010 Oct;25(10):1024-9. PMID: 20480250.

9. Bertholet N, **Cheng DM**, **Samet JH**, Quinn E, **Saitz R**. Alcohol consumption patterns in HIV-infected adults with alcohol problems. Drug Alcohol Depend. Jun 2010;112:160-63. PMID: 20579819.

10. Bickmore TW, Mitchell SE, Jack BW, **Paasche-Orlow MK**, Pfeifer LM, O'Donnell J. Response to a relational agent by hospital patients with depressive symptoms. Interact Comput. 2010 Jul;22(4):289-98. PMID: 20628581.

11. Bickmore TW, Pfeifer LM, Byron D, Forsythe S, **Henault LE**, Jack BW, Silliman R, **Paasche-Orlow MK**. Usability of conversational agents by patients with inadequate health literacy: Evidence from two clinical trials. Journal of Health Communication. 2010;15:197-210. PMID: 20845204.

12. Blood EA, Cabral H, Heeren T, **Cheng DM**. Performance of mixed effects models in the analysis of mediated longitudinal data. BMC Med Res Methodol. 2010 Feb;10:16. PMID: 20170503.

13. **Borzecki AM**, Christiansen CL, Chew P, Loveland S, Rosen AK. Comparison of in-hospital versus 30-day mortality assessments for selected medical conditions. Med Care. 2010 Dec;48(12):1117-21. PMID: 20978451.

14. **Borzecki AM**, Christiansen CL, Loveland S, Chew P, Rosen AK. Trends in the inpatient quality indicators: The Veterans Health Administration experience. Med Care. 2010 Aug;48(8):694-702. PMID: 20613657.

15. **Borzecki AM**, Kader B, **Berlowitz DR**. The epidemiology and management of severe hypertension. Journal Human Hypertension. 2010;24:9-18. PMID: 19440209.

16. Brooks EL, Rosner Preis S, Hwang S-J, **Murabito JM**, Benjamin EJ, Kelly-Hayes M, Sorlie P, Levy D. Health insurance and cardiovascular disease risk factors. Am J of Med. 2010;123(8):741-47. PMID: 20670729. PMCID: PMC2913281.

17. **Chen DCR,** Pahilan ME, **Orlander JD**. Comparing a self-administered measure of empathy with observed behavior among medical students. J of Gen Intern Med. Mar 2010;25(3):200-2. PMID: 20013070.

18. Conrad P, **Carr P**, Knight S, Renfrew MR, Dunn MB, Pololi L. Hierarchy as a barrier to advancement for women in academic medicine. J Womens Health (Larchmt). 2010 Apr;19(4):799-805. PMID: 20141385.

19. Cook ED, Arnold KB, **Hermos JA**, McCaskill-Stevens W, Moody-Thomas S, Probstfield JL, Hamilton SJ, Campbell RD, Anderson KB, Minasian LM. Impact of supplemental site grants to increase African-American accrual for Selenium and Vitamin E Cancer Prevention Trial (SELECT). Clin Trials. 2010;7(1):90-9. PMID: 20156960.

20. Cook RL, McGinnis KA, **Samet JH**, Fiellin DA, Rodriquez-Barradas MC, Kraemer KL, Gibert CL, Braithwaite RS, Goulet JL, Mattocks K, Crystal S, Gordon AJ, Oursler KK, Justice AC. Erectile dysfunction drug receipt, risky sexual behavior and sexually transmitted diseases in HIV-infected and HIV-uninfected men. J Gen Intern Med. 2010 Feb;25(2):115-21. PMID: 19921112.

21. Cortés DE, Drainoni M-L, **Henault LE**, **Paasche-Orlow MK**. How to achieve informed consent for research from Spanish-speaking individuals with low literacy: A qualitative report. J of Health Communication. 2010;15:172-82. PMID: 20845202.

22. Cowan A, **Hylek EM**. Bleeding risk on warfarin among elderly patients with atrial fibrillation. J Am Coll Cardiol. 2010 Mar 2;55(9):932; author reply 932-3. PMID: 20185052.

23. **Crosby SS**, Mohan S, DiLoreto C, Spiegel JH. Head and neck sequelae of torture. Laryngoscope. 2010 Feb;120(2):414-9. PMID: 19998347.

24. Dookeran NM, **Battaglia T**, Cochran J, Geltman PL. Chronic disease and its risk factors among refugees and asylees in Massachusetts, 2001-2005. Prev Chronic Dis. 2010 May;7(3):A51. PMID: 20394690.

25. **Dupuis EA**, White HF, **Newman D**, Sobieraj JE, Gokhale M, **Freund KM**. Tracking abnormal cervical cancer screening: Evaluation of an EMR-based intervention. J Gen Intern Med. 2010 Jun;25(6):575-80. PMID: 20204536.

26. Elder RW, Lawrence B, Ferguson A, **Naimi TS**, Brewer RD, Chattopadhyay SK, Toomey TL, Fielding JE, et al. The effectiveness of tax policy interventions for reducing excessive alcohol consumption and related harms. Am J Prev Med. 2010;38(2):217-229. PMID: 20117579.

27. Elks CE, **Murabito JM**, et al. Thirty new loci for age at menarche identified by a meta-analysis of genome-wide association studies. Nature Genetics. 2010;42(12):1077–85. doi:10.1038/ng.714. PMID: 21102462. PMCID: PMC3140055.

28. Elliott KS, Zeggini E, McCarthy MI, Gudmundsson J, Sulem P, Stacey SN, Thorlacius S, Amundadottir L, Grönberg H, Xu J, Gaborieau V, Eeles RA, Neal DE, Donovan JL, Hamdy FC, Muir K, Hwang S-J, Spitz MR, Zanke B, Carvajal-Carmona L, Brown KM; Australian Melanoma Family Study Investigators, Hayward NK, Macgregor S, Tomlinson IP, Lemire M, Amos CI, **Murabito JM**, Isaacs WB, Easton DF, Brennan P; PanScan Consortium, Barkardottir RB, Gudbjartsson DF, Rafnar T, Hunter DJ, Chanock SJ, Stefansson K, Ioannidis JP. Evaluation of association of HNF1B variants with diverse cancers: Collaborative analysis of data from 19 genome-wide association studies. PLoS One. 2010 May;5(5):e10858. PMID: 20526366. PMCID: PMC2878330.

29. Fox CS, MD, Massaro JM, Schlett CL, Lehman SJ, Meigs JB, O'Donnell CJ, Hoffmann U, **Murabito JM**. Periaortic fat deposition is associated with peripheral arterial disease: The Framingham Heart Study. Circ Cardiovasc Imagin. 2010;3:515-19. doi: 10.1161/CIRCIMAGING.110.958884. PMID: 20639302. PMCID: PMC3060043.

30. Field CA, Baird J, **Saitz R**, Caetano R, Monti PM. The mixed evidence for brief intervention in emergency departments, trauma care centers, and inpatient hospital settings: What should we do? Alcohol Clin Exp Res 2010; 34(12):2004-10. doi: 10.1111/j.1530-0277.2010.01297.x. PMID: 20860610.

31. Frayne SM, Berg E, Holmes TH, Laungani K, **Berlowitz DR**, Miller DR, Pogach L, Jackson VW, Moos R. Mental illness-related disparities in length of stay: Algorithm choice influences results. J Rehab Res Dev. 2010;47(8):709-18. PMID: 21110246.

32. Frayne SM, Miller DR, Sharkansky EJ, Jackson VW, Wang F, Halanych JH, **Berlowitz DR**, Kader B, Rosen CS, Keane TM. Using administrative data to identify mental illness: What approach is best? Am J Med Qual. 2010;25(1):42-50. PMID: 19855046.

33. Freiberg MS, McGinnis KA, Kraemer K, **Samet JH**, Conigliaro J, Ellison RC, Bryant K, Kuller LH, Justice AC; VACS Project Team. The association between alcohol consumption and prevalent cardiovascular diseases among HIV-infected and HIV-uninfected men. J Acquir Immune Defic Syndr. 2010 Feb;53(2):247-53. PMID: 20009766.

34. Garcia DA, Lopes RD, **Hylek EM**. New-onset atrial fibrillation and warfarin initiation: High risk periods and implications for new antithrombotic drugs. Thromb Haemost. 2010 Dec;104(6):1099-1105. PMID: 20886196. doi:10.1160/TH10-07-0491.

35. Garcia FA, **Freund KM**, Berlin M, Digre KB, Dudley DJ, Fife RS, Gabeau G, Geller SE, Magnus JH, Trott JA, White HF. Progress and priorities in the health of women and girls: a decade of advances and challenges. J Womens Health (Larchmt). 2010 Apr;19(4):671-80. PMID: 20201706.

36. Givens JL, Tjia J, Zhou C, Emanuel E, **Ash AS**. Racial and ethnic differences in hospice use among patients with heart failure. Arch Intern Med. 2010 Mar;170(5):427-32. PMID: 20212178.

37. Hahn RA, Kuzara JL, Elder R, Brewer R, Chattopadhyay S, Fielding J, **Naimi TS**, Toomey T, Middleton JC, Lawrence B; Task Force on Community Preventive Services. Effectiveness of policies restricting hours of alcohol sales in preventing excessive alcohol consumption and related harms. Am J Prev Med. 2010 Dec;39(6):590-604. PMID: 21084080.

38. Hahn JA, **Samet JH**. Alcohol and HIV disease progression: Weighing the evidence. Curr HIV/AIDS Rep. 2010;7:226-33. PMID: 20814765.

39. **Hanchate AD**, Clough-Gorr KM, **Ash AS**, Thwin SS, Silliman RA. Longitudinal patterns in survival, comorbidity, healthcare utilization and quality of care among older women following breast cancer diagnosis. J Gen Intern Med. 2010;25(10):1045-50. PMID: 20532657. doi: 10.1007/s11606-010-1407-9.

40. **Hanchate AD**, Stukel TA, Birkmeyer JD, **Ash AS**. Surgery volume, quality of care and operative mortality in coronary artery bypass graft surgery: A re-examination using fixed-effects regression. Health Serv Outcomes Res Method. 2010;10(1-2):16-32. doi: 10.1007/s10742-010-0063-1.

41. Hausmann LRM, **Kressin NR**, Hanusa BH, Ibrahim SA. Perceived racial discrimination in health care and its association with patients' healthcare experiences: Does the measure matter? Ethnicity and Disease. 2010;20(1):40-7. PMID: 20178181.

42. Howland J, Rohsenow DJ, Greece JA, Littlefiled CA, **Almeida A**, Heeren T, Winter M, Bliss CA, Hunt S, **Hermos J**. The effects of binge drinking on college students' next-day academic test-taking performance and mood state. Addiction. 2010 Apr;105(4):655-65. doi:10.1111/j.1360-0443.2009.02880.x.

43. **Hylek EM**. Therapeutic potential of oral factor Xa inhibitors. N Engl J Med. 2010 Dec;363(26):2559-61. PMID: 21175319.

44. **Jackson AH**, **Alford DP**, Dubé CE, **Saitz R**. Internal medicine residency training for unhealthy alcohol and other drug use: Recommendations for curriculum design. BMC Med Educ. 2010 Mar;10:22. PMID: 20230607.

45. Johansson I, Holgerson PL, **Kressin NR**, Nunn ME, Tanner AC. Snacking habits and caries in young children. Caries Res. 2010;44(5):421-30. PMID: 20720422.

46. Kanasi E, Johansson I, Lu SC, **Kressin NR**, Nunn ME, Kent R Jr., Tanner ACR. Microbial risk markers for childhood caries in pediatricians' offices. Journal of Dental Research. 2010 Apr;89(4):378-83. PMID: 20164496.

47. **Kapoor A**, Chuang W, **Radhakrishnan N**, Smith KJ, **Berlowitz D**, Segal JB, Katz JN, Losina E. Cost effectiveness of venous thromboembolism pharmacological prophylaxis in total hip and knee replacement: A systematic review. Pharmacoeconomics. 2010 Jul;28(7):521-38. doi: 10.2165/11535210-000000000-00000. PMID: 20550220.

48. **Kapoor A**, Labonte AJ, Winter MR, Segal JB, Silliman RA, Katz JN, Losina E, **Berlowitz D**. Risk of venous thromboembolism after total hip and knee replacement in older adults with comorbidity and co-occurring comorbidities in the Nationwide Inpatient Sample (2003-2006). BMC Geriatr. 2010 Sep;10:63. PMID:20846450.

49. Krasnoff JB, Basaria S, Pencina MJ, Jasuja GK, Vasan RS, Ulloor J, Zhang A, Coviello A, Kelly-Hayes M, D'Agostino RB, Wolf PA, Bhasin S, **Murabito JM**. Free testosterone levels are associated with mobility limitation and physical performance in community-dwelling men: The Framingham Offspring Study. J Clin Endocrin Metab. 2010 Jun;95(6):2790-9. doi:10.1210/jc.2009-2680. PMID: 20382680. PMCID: PMC2902069.

50. **Kressin NR**, **Manze M**, Russell SL, Katz RV, Claudio C, Green BL, Wang MQ. Self-reported willingness to have cancer screening and the effects of sociodemographic factors. J Natl Med Assoc. 2010 Mar;102(3):219-27. PMID: 20355351.

51. **Kressin NR**, Orner MB, **Manze M**, Glickman ME, **Berlowitz D**. Understanding contributors to racial disparities in blood pressure control. Circ Cardiovasc Qual Outcomes. 2010 Mar;3(2):173-80. PMID: 20233981.

52. **Kronman AC**, **Freund KM**, **Hanchate A**, Emanuel EJ, **Ash AS**. Nursing home residence confounds gender differences in Medicare utilization: An example of Simpson's Paradox. Women's Health Issues. 2010;20:105-13. PMID: 20149970.

53. Levy D, Splansky GL, Strand NK, Atwood LD, Benjamin EJ, Blease S, Cupples LA, D'Agostino RB Sr, Fox CS, Kelly-Hayes M, Koski G, Larson MG, Mutalik KM, Oberacker E, O'Donnell CJ, Sutherland P, Valentino M, Vasan RS, Wolf PA, **Murabito JM**. Consent for genetic research in the Framingham Heart Study. Am J Med Genet A. 2010 May;152A(5):1250-6. PMID: 20425830. PMCID: PMC2923558.

54. **Liebschutz JM**, **Saitz R**, Weiss RD, Averbuch T, Schwartz S, **Meltzer EC**, Claggett-Borne E, Cabral H, **Samet JH**. Clinical factors associated with prescription drug use disorder in urban primary care patients with chronic pain. J Pain. 2010;11(11):1047-55. PMID: 20338815.

55. **Liebschutz J**, Schwartz S, **Hoyte J**, Conoscenti L, Christian AB Sr., Muhummad L, Harper D, James T. Chasm between injury and care: Experiences of black male victims of violence. J Trauma. 2010 Dec;69(6):1372-8. PMID: 20838259.

56. **Linsky A**, Gupta K, Lawler EV, Fonda JR, **Hermos JA**. Proton pump inhibitors and risk for recurrent *Clostridium difficile* infection. Arch Intern Med. 2010 May;170(9):772-8. PMID: 20458084.

57. 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. Am Heart J. 2010 Mar;159(3):331-9. Erratum in: Am Heart J. 2010 Mar;159(3):331-9. PMID: 20211292.

58. Loucks EB, Pilote L, Lynch JW, Richard H, Almeida ND, Benjamin EJ, **Murabito JM**. Life course socioeconomic position is associated with inflammatory markers: The Framingham Offspring Study. Soc Sci Med. 2010 Jul;71(1):187-95. PMID: 20430502. PMCID: PMC2895737.

Update in Pain Medicine

Daniel P. Alford, MD, MPH[1], Erin E. Krebs, MD, MPH[2], Ian A. Chen, MD, MPH[3], Christina Nicolaidis, MD, MPH[4], Matthew J. Bair, MD, MS[2], and Jane Liebschutz, MD, MPH[1,5]

[1] Boston University School of Medicine and Boston Medical Center, Boston, MA, USA; [2] VA Center on Implementing Evidence-Based Practice and Indiana University School of Medicine, Indianapolis, IN, USA; [3] Eastern Virginia Medical School, Norfolk, VA, USA; [4] Oregon Health and Science University, Portland, OR, USA; [5] Boston University School of Public Health, Boston, MA, USA.

KEY WORDS: primary care; chronic pain; analgesics; opioids; complementary and alternative medicine.
J Gen Intern Med 25(11):1222–6
DOI: 10.1007/s11606-010-1452-4

INTRODUCTION

More than 75 million Americans have chronic pain.[1] Pain accounts for 20% of all outpatient visits[2] and over $100 billion dollars per year in direct and indirect costs;[3] analgesics account for 12% of all prescriptions.[4]Our aims were to review recent pain medicine studies and their key findings, and to discuss the implications of these findings for generalist clinical practice.

We systematically searched from January 1, 2008 through December 31, 2009 for peer-reviewed articles that could potentially change generalist care of patients with chronic pain. We searched MEDLINE and PubMed using the medical subject heading (MeSH) terms Pain, Arthritis, Fibromyalgia, Headache Disorders, Pain Measurement, Analgesics or *Narcotics*, and keywords for chronic, persistent, noncancer or primary care, excluding acute pain, postoperative pain, cancer pain, chest pain and pediatrics, and limiting to humans, English language and study type (trial, epidemiologic, review, meta-analysis or guideline). Members of the Society of General Internal Medicine's Pain Medicine Interest Group also suggested relevant articles. The search produced 1,051 references that were narrowed down to 47 based on relevance to generalist practice. Using a 5-point Likert scale, we independently rated these articles on impact on general internal medicine clinical practice, clinical policy and research and quality of study methods. We selected ten articles with the highest ratings.

ASSESSMENT

Krebs EE, Lorenz KA, Bair MJ et al. Development and initial validation of the PEG, a three-item scale assessing pain intensity and interference. J Gen Intern Med. 2009; 24 (6):733–738.

Received June 3, 2010
Revised June 18, 2010
Accepted June 30, 2010
Published online July 15, 2010

Single-item pain assessments do not adequately measure chronic pain, yet multi-dimensional pain scales such as the Brief Pain Inventory (BPI) are impractical for use in primary care. This study developed an ultra-brief pain measure derived from the BPI. Krebs et al. conducted a secondary data analysis of two primary care studies: a longitudinal study of 500 patients with chronic pain and a cross-sectional study of 646 veterans.

The authors assessed reliability (Cronbach's alpha), construct validity (Pearson's correlation coefficients) relative to other measures of pain and function, and responsiveness to change. The resulting scale consists of three items ("Pain average," "interference with Enjoyment of life" and "interference with General activity," or PEG). The PEG demonstrated good internal consistency (alpha 0.73–0.89), construct validity (r=0.6–0.95) and responsiveness to change.

Implications for Practice

The PEG is an appealing instrument to use in primary care because of its brevity and assessment of multiple domains of pain. This study suggests that the PEG has good psychometric properties, and its responsiveness to change is equivalent to the full BPI, from which it is derived. Replacing the commonly used single-item 0–10 Numeric Rating Scale with the three-item PEG may be a desirable strategy in primary care settings. The incremental "cost" of asking three items instead of one appears to be small relative to the potential gain in a more comprehensive understanding of patients' pain experiences.

COLLABORATIVE CARE

Becker A, Leonhardt C, Kochen MM, et al. Effects of two guideline implementation strategies on patient outcomes in primary care: a cluster randomized controlled trial. Spine. 2008;33(5): 473–480.

Although high-quality low back pain (LBP) guidelines are widely available, evidence-based LBP management strategies are inconsistently applied in primary care. The best methods of LBP guideline implementation are unclear. This trial randomized 118 German primary care practices to one of two LBP guideline implementation strategies or to a mailed guidelines control. The implementation strategies included physician education (three seminars and two academic detailing sessions) alone or physician education plus nurse training in motivational counseling (two full-day seminars and one to

three supervised practice sessions). Participating physicians enrolled 1,278 patients (fewer than the enrollment target of 1,874).

Most participants were male (58%) and employed (55%). Compared with those in the control group, patients in the nurse training group improved significantly more on the primary outcome of functional capacity at 6 months (p=0.032). There was no difference between the physician education only and control groups (p=0.120). Intervention patients reported fewer days in pain than control patients (16.4 and 17.9 fewer days).

Implications for Practice

Intervention effects were small, but the results have significant implications for generalist practice. First, the study enrolled a broad group of patients with LBP, of whom <50% had pain at follow-up. Many patients had resolution of their pain, which may partially explain the modest intervention effects on pain-related function. Second, the intervention was relatively non-intensive and was delivered distal to the clinical encounter. Although this likely reduced the intervention effect, implementation of a similar intervention in actual practice would be more feasible. Considering these factors, the finding of a small improvement in patient function is impressive.

The trial's finding that training both clinic nurses and physicians improved patient function, whereas training physicians alone did not, supports the concept of team-based primary care for back pain. Although some guideline-recommended practices can be implemented by physicians alone (e.g., limiting imaging tests), others may be challenging for physicians to implement without support (e.g., counseling for increased physical activity).[5] Nurses may be better equipped to provide patient education and counseling interventions.

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.

Collaborative interventions based on the chronic care model have been studied in numerous conditions,[6] but have not been rigorously evaluated for chronic pain management. Dobscha et al. evaluated a multifaceted collaborative care intervention for chronic musculoskeletal pain in five Veterans Affairs (VA) primary care clinics. They randomized 42 primary care clinicians to the collaborative care intervention or usual care control. The intervention included pain management training for clinicians, pain education classes for patients, and care management by a full-time psychologist who assessed patients, developed tailored care recommendations collaboratively with a pain consultant and followed up with patients by phone. Patients (n=401) with chronic musculoskeletal pain of at least moderate pain severity and disability were enrolled.

Most participants were male (92%) and had long-standing pain (mean=10 years); 32% were employed. The mean number of contacts with the collaborative care team was ten, including an average of five phone contacts. On the primary outcome (pain-related disability) at 12 months, intervention patients improved significantly more than controls (p=0.004). A higher proportion of intervention patients experienced clinically important improvements of ≥30% than controls (22% vs. 14%, p=0.04, NNT=12.7). Secondary outcomes of pain intensity and depression severity were also improved in the intervention group. The intervention cost $1,200 per patient.

Implications for Practice

This trial showed that a collaborative primary care-based intervention can lead to modest, yet important improvements in pain outcomes among highly disabled, long-term pain sufferers. The collaborative care team directly intervened with patients by clinically assessing and reassessing them, delivering education to them, and providing their clinicians with individualized treatment recommendations. Even then, the intervention was relatively inexpensive, as it involved only one additional full-time clinician (the psychologist care manager) and 10% of an internist consultant's time. Considering the amount of health care consumed by patients with severe chronic pain,[7] an investment in collaborative care could be a bargain.

Clinician satisfaction, an important outcome, was not reported in this study. In a previous paper, Dobscha et al. reported three-fourths of VA primary care clinicians rated chronic pain management as a major source of frustration.[8] Improving satisfaction of primary care providers is another compelling reason to implement collaborative care interventions that support both patients and providers in addressing chronic pain.

CHRONIC PAIN AND DEPRESSION

Kroenke K, Bair MJ, Damush TM, et al. Optimized antidepressant therapy and pain self-management in primary care patients with depression and musculoskeletal pain: a randomized controlled trial. JAMA. 2009;301(20): 2099–2110.

Pain and depression frequently coexist (30%–50% co-occurrence), leading to poor health-related quality of life (HRQoL).[9,10] This study evaluated a stepped-care intervention including optimized antidepressant therapy combined with a pain self-management program to improve pain and depression outcomes. The study randomized 250 patients with chronic low back, hip, or knee pain of moderate severity and depression of moderate severity to intervention versus usual care. At 12 months, intervention patients were more likely than controls to experience a 50% reduction from baseline in depression severity (37.4% versus 16.5%, RR 2.3, p<0.001). Intervention patients also showed statistically significant improvements in pain severity and interference and in secondary outcomes (anxiety and HRQoL).

Implications for Practice

This trial demonstrated that optimized antidepressant therapy combined with a pain self-management program was beneficial for treating patients with co-occurring depression and pain. Improvements in depression and pain were seen early and sustained over 12 months. The improvements in pain severity and interference are noteworthy since analgesics were not a specific intervention component. The optimized antidepressant therapy included an algorithmic approach using several classes of antidepressants. This pragmatic, patient-tailored approach is more similar to clinical practice than an

inflexible testing of a single drug or class. Practitioners should screen for co-morbid depression and pain in their patients, and consider optimized antidepressant therapy and self-management strategies as essential tools in the treatment of these patients.

OPIOIDS AND CHRONIC PAIN

Weisner CM, Campbell CI, Ray GT, et al. Trends in prescribed opioid therapy for non-cancer pain for individuals with prior substance use disorders. Pain 2009;145(3): 287–293.

The efficacy and risks of long-term opioids in patients with substance abuse histories are poorly understood and prescribing practices not well described. Experts recognize that patients with substance use disorders commonly have pain that needs to be treated, but advise caution when prescribing opioids in this context. This study examined trends and characteristics of long-term opioid use in patients with non-cancer pain and substance use disorders (SUD). Administrative data from two large community health plans (Kaiser Permanente of Northern California and Group Health Cooperative of Seattle Washington) were evaluated. Long-term opioid use was defined as treatment >90 days. From 1997–2005, opioid prescribing increased dramatically for patients with SUD diagnosed prior to opioid prescription and was highest for those with an opioid use disorder. Patients with a SUD history were prescribed significantly more opioids (dose, potency and supply) and more concurrent sedative-hypnotics compared to patients without SUD histories. Additional findings were that more than a third of patients with a SUD history and almost three-fourths with a prior opioid use disorder received these diagnoses from addiction or psychiatric providers, rather than from the prescriber of the long-term opioid.

Implications for Practice

The increasing opioid prescribing for patients with SUD reported in this study are of unclear clinical appropriateness. Experimental studies[11] have shown higher pain sensitivity among these patients, and observational studies[12] have shown they have a greater risk of prescription opioid misuse. This study did not examine whether opioid prescribers were aware of their patients' substance use histories. At a minimum, generalists should screen all patients with chronic pain for a history of substance abuse before prescribing long-term opioids. Single-item screening[13,14] tests for at-risk substance use are now available. Patient monitoring with "universal precautions[15]," including opioid treatment agreements and urine drug testing, are recommended for all patients prescribed long-term opioids for chronic pain. The level of monitoring may be reasonably intensified (e.g., pill counts and random urine drug testing) for patients with a history of substance use disorders. In addition, based on the high rate of concurrent sedative use, generalists should carefully educate these patients about the risk of over-sedation.

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):113–130.

A multidisciplinary expert panel, commissioned by the American Pain Society and the American Academy of Pain Medicine, conducted a systematic review of the evidence and developed evidence-based guidelines for use of chronic opioid therapy in chronic non-cancer pain. They reviewed the literature through November 2007, including 8,034 relevant abstracts. The expert panel met three times between September 2006 and January 2008, using methods adapted from the GRADE working group and multi-stage Delphi process. The panel found that evidence was limited, with many "research gaps," but concluded that chronic opioids "can be an effective therapy for carefully selected and monitored patients with chronic non-cancer pain." Recommendations were provided on topics including: patient selection, medication management plans, monitoring strategies, prevention and management of adverse effects, and use of psychotherapeutic co-interventions.

Implications for Practice

These recommendations provide a reasonable and comprehensive guide for general internists when prescribing long-term opioids for chronic pain. As the authors acknowledge, many of their recommendations are limited by reliance on low-quality evidence. The Chou et al. companion paper on research gaps in opioid use provides important background to the guidelines.[16] Recommendations related to assessing substance abuse risk, using informed consents and monitoring strategies, including urine drug testing for "high risk" patients, and counseling patients about driving risk are likely to become accepted "standard of care" until research emerges to better guide clinical practice.

CANNABIS AND CHRONIC PAIN

Martin-Sanchez E, Furukawa TA, Taylor J, Martin JLR. Systematic review and meta-analysis of cannabis treatment for chronic pain. Pain Med 2009;10(8):1353–1368.

Cannabis has been used for the treatment of various conditions, including chronic pain; however, little is known about its effectiveness or harms. The authors conducted a systematic review and meta-analysis of double-blind randomized controlled trials comparing cannabis preparations to placebo in patients with chronic pain. Cannabis preparations varied in dose, duration and route of administration (oral or nasal). The authors excluded one study of smoked cannabis on "ethical grounds." Of 229 studies, 18 met the inclusion criteria, but only 7 trials could be quantitatively analyzed because of data reporting limitations.

Cannabis reduced pain intensity compared to placebo, with an overall effect size of -0.61 (-0.84 to -0.37). The identified harms included euphoria (OR 4.11; NNH 8), dysphoria (OR 2.56; NNH 29), disturbances in perception (OR 4.5; NNH 7) and motor function (speech, ataxia, twitching, numbness; OR 3.9; NNH 5), altered cognitive function (OR 4.4; NNH 8) and gastrointestinal side effects (risks noted, but no summary due to heterogeneity.)

Implications for Practice

This study summarizes the evidence for benefits and harms of cannabis-derived compounds in the treatment of chronic pain. Notable limitations of the existing literature include small

sample sizes (n=13–177), short follow-up periods (mean=25 days), incomplete outcome reporting and exclusion of studies of smoked cannabis. Benefits and risks do not seem very dissimilar from other pharmacologic treatments (e.g., opioids) for chronic pain. This study offers preliminary evidence to inform the complex clinical decisions of risk-benefit for individual patients. Clearly, more rigorous studies are needed to better understand the risks and benefits of cannabis preparations for treating chronic pain.

COMPLEMENTARY AND ALTERNATIVE TREATMENTS

Cherkin DC, Sherman KJ, Avins AL, et al. A randomized trial comparing acupuncture, simulated acupuncture, and usual care for chronic low back pain. Arch Intern Med. 2009;169(9):858–866.

Back pain is the leading reason for visits to acupuncturists.[17] Several recent European trials have suggested that real acupuncture and "sham" acupuncture are equally effective.[18–20] This four-arm randomized controlled trial compared individually tailored acupuncture, standard acupuncture, simulated acupuncture (non-insertive) and usual care for 638 adults with mechanical low back pain. Detailed blinding procedures prevented acupuncture patients from knowing their treatment assignment. At 8 weeks, all three acupuncture arms had significantly lower pain disability scores than the usual care arm. Clinically meaningful improvements occurred in 60% of acupuncture patients compared with 30% in usual care (P<0.001). Improvements persisted at 1 year. There were no differences between individually tailored, standard and simulated acupuncture.

Implication for Practice

Acupuncture appears to be effective for improving back pain-related disability. Interestingly, insertion of needles and individual tailoring by acupuncturists did not make a difference in outcomes. For practitioners, the mechanism for effectiveness may not be as relevant as having an effective non-pharmacological therapy with minimal risks in the armamentarium of treatment options for the low back pain. Clinical trials evaluating the efficacy of acupuncture for various conditions have been conducted; however, most were poorly designed, and few included treatment protocols that can be applied in actual clinical practice. This study was rigorously conducted and raises questions about the putative mechanism of action for acupuncture's effectiveness. Similarly, although exercise for back pain has been shown to improve work disability,[21] it doesn't appear that the type of exercise itself makes the difference.[22] Thus, physicians should encourage patients who express interest in acupuncture to pursue it as a reasonable option.

TREATMENT FOR SPECIFIC TYPES OF CHRONIC PAIN

Chou R, Carson S, Chan BK. Gabapentin versus tricyclic antidepressants for diabetic neuropathy and post-herpetic neuralgia: discrepancies between direct and indirect meta-analyses of randomized controlled trials. J Gen Intern Med. 2009;24(2):178–188.

Tricyclic antidepressants (TCA) and gabapentin have been recommended as first-line treatments for neuropathic pain.[23] Previous systemic reviews have suggested that TCAs were superior to gabapentin, but head-to-head trials were not included in those analyses. Additionally, the quality and interpretability of neuropathic pain trials have been questioned because many of the placebo-controlled trials were published in different time periods (TCAs before 1991 and gabapentin after 1998).

The authors performed two meta-analyses that evaluated the effects of gabapentin versus TCAs in direct (head-to-head) and indirect comparisons (each drug versus placebo). Included trials were rated for quality (randomization, allocation concealment, blinding and intention-to-treat analysis). Gabapentin was superior to placebo for pain relief in six trials (RR 2.18, 1.78–2.67, p<0.00001), and TCAs were also superior to placebo for pain relief in nine trials (RR 5.27, 3.05–9.11, p<0.0001). Indirect comparisons of these 15 trials suggest gabapentin to be inferior to TCAs (RR 0.41, 0.23–0.74), but the direct, head-to-head comparisons (n=3 trials) showed no difference (RR=0.99, 0.76–1.29).

Implications for Practice

This study suggests that gabapentin and TCAs do not differ in efficacy when compared head-to-head, but that TCAs are slightly superior to gabapentin in indirect comparisons. For several reasons, trial results should be interpreted with caution. First, the placebo response rates were considerably different between gabapentin (24%) and TCAs (6%). Second, sample sizes were markedly different between trials (gabapentin median 112 and TCAs median 26). Third, none of the TCA trials met methodological quality standards. Lastly, none of the trials for either drug lasted longer than 12 weeks. Clinicians should be aware of the possible differences in outcomes due to quality and methodological variations in trials. The treatment of neuropathic pain with either gabapentin or TCAs is reasonable and should depend on factors such as patient preference, cost or side effects.

Hauser W, Bernardy K, Uceyler N, Sommer C. Treatment of fibromyalgia syndrome with antidepressants: a meta-analysis. JAMA. 2009;301(2):198–209.

Although prior meta-analyses have shown tricyclic antidepressants (TCA) to be effective in reducing pain in fibromyalgia syndrome (FMS),[24,25] newer classes of anti-depressants have not been evaluated. This meta-analysis reviewed 18 randomized controlled trials (2,296 total patients) comparing antidepressant with pharmacological placebo for treatment of FMS. Outcomes included standardized measures of pain, fatigue, sleep quality, depressed mood and HRQoL. The authors compared antidepressant classes as well as individual antidepressants.

Overall, antidepressants significantly improved pain [effect size (ES)=-0.43, p<0.001)], depressed mood (ES=-0.26, p<0.001), sleep (ES=-0.32, p<0.001) and HRQoL (ES=-0.31, p<0.001), but had no effect on fatigue. TCAs showed large effects for pain, fatigue and sleep, a small effect for HRQoL and no effect for depressed mood. Selective serotonin reuptake inhibitors (SSRI) showed a small effect for pain, depressed mood and HRQoL, but no effect on fatigue or sleep. Serotonin-norepinephrine reuptake inhibitors (SNRI) showed small effects for pain, sleep,

depressed mood and HRQoL, and no effect on fatigue. MAO inhibitors showed a moderate effect for pain, but none for fatigue, sleep or depressed mood. Amitriptyline and duloxetine had the strongest evidence for symptom improvement of the individual medications.

Implications for Practice

When considering pharmacotherapy for FMS, antidepressants should be among the first choices. Among the antidepressant choices, TCAs, particularly amitriptyline, has consistently been shown to have a strong effect on many FMS symptoms, an effect that has been stronger than with SSRIs and SNRIs, albeit with a higher rate of side effects. Providers should educate patients to expect only small to moderate improvement in symptoms that characterize FMS—pain, sleep disturbance and depressed mood—as well as in HRQoL. Fatigue is unlikely to improve with antidepressants. Side effect profiles should be weighed against expected improvement in symptoms. In addition, treatment should include evidence-based non-pharmacological treatments, such as such as tailored exercise, heated pool therapy, cognitive behavioral therapy and relaxation techniques.[26]

Acknowledgements: *All authors are members of the Society of General Internal Medicine's Pain Medicine Interest Group. This paper derives from the presentation "Update in Pain Medicine" at the 33rd annual meeting of SGIM on April 2010 in Minneapolis, Minnesota. The authors thank Joel Hoyte at Boston Medical Center for his administrative support in coordinating the submission of this manuscript. Drs. Bair and Krebs are supported by VA HSR&D Research Career Development Awards.*

Potential Financial Conflicts of Interest: *Dr. Bair has consulted for Abbott and Cephalon. The other authors reported no potential conflicts of interest.*

Corresponding Author: *Daniel P. Alford, MD, MPH; Clinical Addiction Research and Education Unit, Section of General Internal Medicine, Boston Medical Center, 801 Massachusetts Avenue, 2nd Floor, Boston, MA 02118, USA (e-mail: dan.alford@bmc.org).*

REFERENCES

1. National Center for Health Statistics. United States, 2006, with chartbook on trends in the health of Americans. 68–71. 2006. Hyattsville, MD.
2. **Schappert SM.** National Ambulatory Medical Care Survey: 1989 summary. Vital Health Stat. 1992;13:1–80.
3. **Turk DC, Okifuji A, Kalauokalani D.** Clinical outcome and economic evaluation of multidisciplinary pain centers. In: **Block AR, Kremer EF, Fernandez E, eds.** Handbook of pain syndromes: Biopsychosocial perspectives. Mahwah: Erlbaum; 1999:77–98.
4. **Turk DC.** Clinical effectiveness and cost-effectiveness of treatments for patients with chronic pain. Clin J Pain. 2002;18(6):355–65.
5. **Chou R, Qaseem A, Snow V, et al.** Diagnosis and treatment of low back pain: A joint clinical practice guideline from the American College of Physicians and the American Pain Society. Ann Intern Med. 2007;147 (7):478–91.
6. **Bodenheimer T, Wagner EH, Grumbach K.** Improving primary care for patients with chronic illness. The chronic care model, part 2. JAMA. 2002;288(14):1775–9.
7. **Martin BI, Deyo RA, Mirza SK, et al.** Expenditures and health status among adults with back and neck problems. JAMA. 2008;299(6):656–64.
8. **Dobscha SK, Corson K, Flores JA, Tansill EC, Gerrity MS.** Veterans affairs primary care clinicians' attitudes toward chronic pain and correlates of opioid prescribing rates. Pain Med. 2008;9(5):564–71.
9. **Bair MJ, Robinson RL, Katon W, Kroenke K.** Depression and pain comorbidity: A literature review. Arch Intern Med. 2003;163(20):2433–45.
10. **Sternbach RA.** Survey of pain in the United States: The Nuprin Pain Report. Clin J Pain. 1986;2:49–53.
11. **Alford DP, Compton P, Samet JH.** Acute pain management for patients receiving maintenance methadone or buprenorphine therapy. Ann Intern Med. 2006;144(2):127–34.
12. **Ives TJ, Chelminski PR, Hammett-Stabler CA, et al.** Predictors of opioid misuse in patients with chronic pain: A prospective cohort study. BMC Health Serv Res. 2006;6:46.
13. **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.
14. **Smith PC, Schmidt SM, Allensworth-Davies D, Saitz R.** A single question screening test for drug use in primary care. Arch. Intern. Med. 2010;170(13):1155–60.
15. **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.
16. **Chou R, Ballantyne JC, Fanciullo GJ, Fine PG, Miaskowski C.** Research gaps on use of opioids for chronic noncancer pain: Findings from a review of the evidence for an American Pain Society and American Academy of Pain Medicine clinical practice guideline. J Pain. 2009;10 (2):147–59.
17. **Cherkin DC, Deyo RA, Sherman KJ, et al.** Characteristics of visits to licensed acupuncturists, chiropractors, massage therapists, and naturopathic physicians. J Am Board Fam Pract. 2002;15(6):463–72.
18. **Brinkhaus B, Witt CM, Jena S, et al.** Acupuncture in patients with chronic low back pain: A randomized controlled trial. Arch Intern Med. 2006;166(4):450–7.
19. **Haake M, Muller HH, Schade-Brittinger C, et al.** German acupuncture trials (GERAC) for chronic low back pain: Randomized, multicenter, blinded, parallel-group trial with three groups. Arch Intern Med. 2007;167(17):1892–8.
20. **Thomas KJ, MacPherson H, Thorpe L, et al.** Randomised controlled trial of a short course of traditional acupuncture compared with usual care for persistent non-specific low back pain. BMJ. 2006;333 (7569):623.
21. **Oesch P, Kool J, Hagen KB, Bachmann S.** Effectiveness of exercise on work disability in patients with non-acute non-specific low back pain: Systematic review and meta-analysis of randomised controlled trials. J Rehabil Med. 2010;42(3):193–205.
22. **Macedo LG, Smeets RJ, Maher CG, Latimer J, McAuley JH.** Graded activity and graded exposure for persistent nonspecific low back pain: A systematic review. Phys Ther. 2010 Apr 15.
23. **Dworkin RH, O'Connor AB, Backonja M, et al.** Pharmacologic management of neuropathic pain: Evidence-based recommendations. Pain. 2007;132(3):237–51.
24. **Arnold LM, Keck PE Jr, Welge JA.** Antidepressant treatment of fibromyalgia. A meta-analysis and review. Psychosomatics. 2000;41 (2):104–13.
25. **O'Malley PG, Balden E, Tomkins G, Santoro J, Kroenke K, Jackson JL.** Treatment of fibromyalgia with antidepressants: A meta-analysis. J Gen Intern Med. 2000;15(9):659–66.
26. **Carville SF, Arendt-Nielsen S, Bliddal H, et al.** EULAR evidence-based recommendations for the management of fibromyalgia syndrome. Ann Rheum Dis. 2008;67(4):536–41.

Drug and Alcohol REVIEW

APSAD

Drug and Alcohol Review (November 2010), 29, 655–661
DOI: 10.1111/j.1465-3362.2010.00222.x

Do attitudes about unhealthy alcohol and other drug (AOD) use impact primary care professionals' readiness to implement AOD-related preventive care?

MICHAELA B. AMARAL-SABADINI[1], RICHARD SAITZ[2,3] &
MARIA LUCIA O. SOUZA-FORMIGONI[1]

[1]*Department of Psychobiology, Federal University of Sao Paulo, Sao Paulo, Brazil,* [2]*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, USA, and* [3]*Department of Epidemiology, Boston University School of Public Health, Boston, USA*

Abstract

Introduction and Aims. *To explore the association between primary care professionals' (PCPs) attitudes towards unhealthy alcohol and other drug (AOD) use (from risky use through dependence) and readiness to implement AOD-related preventive care.* ***Design and Methods.*** *Primary care professionals from five health centres in Sao Paulo were invited to complete a questionnaire about preventive care and attitudes about people with unhealthy AOD use. Logistic regression models tested the association between professional satisfaction and readiness. Multiple Correspondence Analysis assessed associations between stigmatising attitudes and readiness.* ***Results.*** *Of 160 PCPs surveyed, 96 (60%) completed the questionnaire. Only 25% reported implementing unhealthy AOD use clinical prevention practices; and 53% did not feel ready to implement such practices. Greater satisfaction when working with people with AOD problems was significantly associated with readiness to implement AOD-related preventive care. In Multiple Correspondence Analysis two groups emerged: (i) PCPs ready to work with people with unhealthy AOD use, who attributed to such patients lower levels of dangerousness, blame for their condition and need for segregation from the community (suggesting less stigmatising attitudes); and (ii) PCPs not ready to work with people with unhealthy AOD use, who attributed to them higher levels of dangerousness, blame, perceived level of patient control over their condition and segregation (suggesting more stigmatising attitudes).* ***Discussion and Conclusions.*** *More stigmatising attitudes towards people with unhealthy AOD use are associated with less readiness to implement unhealthy AOD-related preventive care. Understanding these issues is likely essential to facilitating implementation of preventive care, such as screening and brief intervention, for unhealthy AOD use.* [Amaral-Sabadini MB, Saitz R, Souza-Formigoni MLO. Do attitudes about unhealthy alcohol and other drug (AOD) use impact primary care professionals' readiness to implement AOD-related preventive care? *Drug Alcohol Rev* 2010;29;655–661]

Key words: alcohol, drug, primary health care, attitude, stigmatisation.

Introduction

The use of alcohol and other drugs (AOD) can be perceived as positive or negative depending on many factors. In many situations alcohol use is associated with prestigious activities and social status. On the other hand, the use of illegal drugs and the way some substances are used seems to attract stigma, marginalisation and negative emotional reactions [1–3].

Michaela B. Amaral-Sabadini PhD, Researcher, Richard Saitz MD, MPH, Professor of Medicine and Epidemiology, Maria Lucia O. Souza-Formigoni PhD, Professor. Correspondence to Dr Maria Lucia O. Souza-Formigoni, UDED, Universidade Federal de São Paulo, Rua Napoleao de Barros, 1038, Vila Clementino, cep 04024-002, São Paulo-SP, Brazil. Tel: +55 (11) 5549 2500; Fax: +55 (11) 5572 5092; E-mail: mlosformigoni@unifesp.br

Received 31 December 2009; accepted for publication 17 June 2010.
Place where the work was undertaken: Department of Psychobiology, Universidade Federal de São Paulo (Federal University of São Paulo—UNIFESP), Brazil.
This study was presented in abstract (poster) form at the INEBRIA Meeting (International Network on Brief Interventions for Alcohol Problems), 8th and 9th October 2009, Newcastle Gateshead (England).

Alcohol and other drug-related attitudes may develop from and impact people's personal and professional experiences and behaviours. Attitude is a hypothetical construct that represents the person's perspective (positive or negative) towards a specified target (person, place, thing or event). Attitudes are generally defined as being composed of three components: cognitive (beliefs), affective (emotions) and behavioural (verbal or typical behavioural tendency). The three components are interlinked and negative beliefs, for example, are associated with stigmatising attitudes [4]. Link and Phelan [5] defined stigma as the co-occurrence of labelling, stereotyping, separation, status loss and discrimination. Weiner [6] applies the attribution theory to understand the perceived cause of stigma. According to this theory, attributions of responsibility will determine affective reactions towards the stigmatised person (e.g. anger or pity), future expectations regarding the individual (e.g. likelihood of recovery) and a variety of behavioural responses, including altruistic actions [7].

Health professionals' negative attitudes, particularly about AOD use, have been cited as important barriers to the implementation of clinical prevention practices [8–10]. Some health professionals consider AOD-related problems difficult issues to discuss, and many report a lack of skills, confidence, satisfaction and competence to identify and manage these problems [11–13]. Ronzani *et al.* [14] pointed out that unhealthy AOD use (the spectrum from risky use through dependence) was the most negatively judged behaviour by health professionals when compared with other health conditions, such as Hansen's disease, obesity, HIV and others.

In order to reach a broad range of people for early intervention, decrease stigma and improve attitudes and care, many have advocated addressing unhealthy AOD use in primary health-care settings [12,15,16], where the focus is on prevention, health promotion and longitudinal comprehensive care. This laudable goal has the potential to make addressing unhealthy AOD use a mainstream health issue and risk behaviour, managed like others addressed routinely in these settings. However, few studies about health professionals' attitudes related to unhealthy AOD use have examined the association between them and the professional's clinical practices. The present study aimed to explore the association between primary care professionals' (PCP) attitudes about people with unhealthy AOD use (including stigma) and PCPs' readiness to implement AOD clinical prevention practices. We anticipated that negative attitudes about people with unhealthy AOD use would influence PCPs' readiness to implement AOD-related preventive care; and that PCPs' satisfaction when working with people with unhealthy AOD use would be associated with readiness to implement AOD-related preventive care.

Methods

Study sample

Physicians, nurses, nursing assistants and community health workers (community health workers are people from the community where the health centre is located, trained and hired to work on community health activities, such as screening and education about health issues) from five primary health-care centres in Sao Paulo, Brazil, were invited to participate in the study. These centres were randomly selected from 222 centres in total, one centre from each of the city's regions—North, South, Southeast, Center-west and East. All the health professionals from each centre were invited to participate ($n = 160$) and 96 completed the survey, giving a response rate of 60%. The participants were physicians or nurses (31%), nursing assistants (19%) and community health workers (50%).

As regards sociodemographic characteristics, 87% were women, 60% single, mean age 40.6 (SD 9.7); 49% were Catholic, 25% Protestant, 24% had other religions or did not have a religion.

Data collection

The first step was to present the study to the primary care unit managers. They were responsible for inviting all the health professionals in the unit to participate. Before agreeing to participate, professionals were told about the aims of the project and informed about the confidentiality of the data (it was an anonymous survey), were asked to sign informed consent and instructed about how to complete the questionnaire. All participants received the questionnaire in an envelope. The researcher waited outside of the room to receive the envelope containing the completed survey, or in some cases, returned to the health centre to pick it up on another day.

Instrument

The instrument was composed of three parts. The first part was a questionnaire developed by the authors, with multiple response options: sociodemographic data, clinical prevention practices, beliefs and satisfaction when working with people with AOD problems and readiness to implement preventive interventions with people with risky AOD use. Readiness was evaluate in two items, one for alcohol, one for drugs ('On a

scale of one to five, where 1 means not even a little ready, and 5 totally ready, how ready do you feel to carry out preventive interventions for risky alcohol/drug use?'). For the questions about beliefs, satisfaction and readiness Likert scales ranging from 1 to 5 were used. AOD-related preventive care was defined as any type of screening or early intervention done in the primary care setting studied. The second questionnaire was an adaptation made by the authors of the Attribution Questionnaire Short Form [3,17], a stigma measure. The Adapted version of the Attribution Questionnaire (AAQ) addressed health professionals' attributions for different health conditions represented by hypothetical vignettes about people with: (i) risky alcohol use; (ii) alcohol abuse; (iii) alcohol dependence; (iv) drug dependence; (v) hypertension; (vi) depression; (vii) schizophrenia; and (viii) HIV-AIDS. Participants were instructed to read the eight vignettes and score nine different items: pity for the patient, dangerousness of the patient, fear of the patient, blame (patient is responsible for his condition), segregation (would be better for the community for the patient to be removed from it for treatment), anger (about or towards the patient), likelihood that the professional would help the patient, avoidance (of taking care of the patient) and control (how much the patients have control over for the solutions of their problems), on a 9-point agreement scale ranging from none/a little (1) to a lot (9). The 'help' item was reverse-coded. Internal consistency (Cronbach's alpha) for each factor was tested: pity: 0.89; dangerousness: 0.77; fear: 0.72; blame: 0.65; segregation: 0.82; anger: 0.77; help: 0.83; avoidance: 0.71; and control: 0.72. We added a final question across the eight vignettes to assess PCPs' opinions about each health condition: 'If you were this person, how would you define your health condition? From 1 (very bad health) to 10 (very good health).' Each vignette was developed considering PCPs' daily practice, including health conditions typically associated with stigma and others, which were not expected to be associated with stigma.

The last part of the instrument was the ASSIST (Alcohol, Smoking and Substance Involvement Screening Test) [18–20] aiming to evaluate the PCPs' AOD use. The ASSIST score categorises respondents as low risk/no use (should receive information), moderate risk (should receive a brief intervention) and high risk (should receive more intensive care).

Statistical analysis

The Pearson χ^2-test was applied to assess the association between categorical independent variables and readiness to implement AOD clinical prevention practices. In logistic regression models, we tested the association between satisfaction when working with people with unhealthy AOD use and readiness to implement AOD-related preventive care. Satisfaction was categorised as 'none' (1 + 2), 'some' (3) and 'a great deal' (4 + 5) and readiness was coded as a binary variable [not ready (1–3) vs. ready (4 + 5)]. $P < 0.05$ was considered statistically significant.

Non-parametric analyses using median regression, with bootstrap estimation were used to compare the participants' opinion about each health condition of the vignettes. Multiple Correspondence Analysis (MCA) was used to assess patterns of associations among the items in the AAQ (medians) and readiness to implement AOD clinical prevention practices (binary). MCA is an exploratory technique (not inferential) that is a generalisation of a principal component analysis for categorical variables. This approach allows testing each difference among groups. Its interpretation can be based upon proximities between points in a low dimensional map, where the axes are the coordinates, the geometric figures represent the different variables in the model and the proximities reflect the strength of associations [21].

Ethics

All the study procedures were approved by the Ethics Committee of the Universidade Federal de São Paulo, project number CEP1581/07.

Results

Over half (56%) of the 96 respondents reported always or almost always implementing general clinical prevention practices, but only 25% reported implementing clinical prevention practices targeting unhealthy AOD use; 53% felt only a little or not at all ready to implement clinical prevention practices for unhealthy AOD use. In spite of that, the majority expressed positive beliefs about working with people with unhealthy AOD use, even though around 60% reported not having adequate training to do so. Sixty-eight per cent of the respondents believed that their amount of work would increase if they start to identify unhealthy AOD use in their clinical routine.

Based on the ASSIST questionnaire, 6.2% of the PCPs used alcohol and 14.6% used tobacco at moderate-risk levels for which receipt of a brief counselling intervention would be appropriate. One used alcohol at a high-risk level. None used drugs. Alcohol and tobacco risk levels were not significantly associated with readiness to implement AOD preventive care for patients with risky use. The proportion ready to implement alcohol preventive care was 24% of those at lower tobacco risk, 43% for moderate risk and 25% of never

Table 1. *Association between: satisfaction and readiness to carry out interventions for risky alcohol use*

		OR (95% CI)
Satisfaction when working with people with unhealthy alcohol use	None	1.0
	Some	4.8 (0.9–26.2)
	A great deal	6.2 (1.6–23.4)
Satisfaction when working with people with unhealthy drug use	None	—
	Some	5.0 (1.0–24.1)
	A great deal	12.0 (3.1–46.6)

CI, confidence interval; OR, odds ratio.

Table 2. *Association between: satisfaction and readiness to carry out interventions for drug use*

		OR (95% CI)
Satisfaction when working with people with unhealthy alcohol use	None	1.0
	Some	1.4 (0.1–16.6)
	A great deal	10.6 (2.2–49.6)
Satisfaction when working with people with unhealthy drug use	None	1.0
	Some	4.0 (0.6–26.3)
	A great deal	18.5 (3.8–89.2)

CI, confidence interval; OR, odds ratio.

smokers; 28% of those at lower alcohol risk, 17% for moderate risk and 25% of never drinkers. The proportion ready to implement drug preventive care was 22% of those at lower tobacco risk, 36% for moderate risk and 25% of never smokers; 25% of those are lower alcohol risk, 17% for moderate risk and 25% of never drinkers.

Greater professional satisfaction when working with people with AOD problems was significantly associated with readiness to implement AOD-related preventive care (Tables 1 and 2). For example, compared with none, having a great deal of professional satisfaction when working with people with alcohol problems increased the odds of readiness to carry out preventive interventions for risky use 6.2 times and the odds of readiness to carry out drug preventive interventions 10.6 times.

We did not find statistically significant differences between professional categories and readiness to implement AOD-related preventive care. For physicians/nurses, nursing assistants and community health workers 37%, 22% and 23%, respectively, were ready to implement alcohol preventive care; similarly, 30%, 22% and 21%, respectively, were ready to implement drug preventive care.

PCPs rated drug dependence, alcohol dependence and alcohol abuse as conditions representing worse health than HIV-AIDS, depression, hypertension and schizophrenia (overall $P < 0.01$ except for schizophrenia) (Figure 1).

The MCA suggested two distinct groups—PCPs that feel ready to carry out preventive interventions for risky AOD use and PCPs that do not (Figure 2). Two patterns of associations between AAQ attitudes and readiness emerged: (i) professionals ready to carry out preventive interventions for risky AOD use, who attributed lower levels of dangerousness, blame (patient is responsible for his condition) and segregation (would be better for the community for the patient to be removed from it for treatment) to such patients (suggesting less stigmatising attitudes); and (ii) professionals not ready to carry out preventive interventions for risky AOD use, who attributed higher levels of dangerousness, blame, control (how much the patient has for the solutions of his problems) and segregation to such patients (suggesting more stigmatising attitudes). Also, within the group of PCPs that do not feel ready, there appear to be two subgroups: those who attribute higher levels of dangerousness and segregation, and those who attribute higher levels of blame and control.

Discussion

The aim of this study was to explore the association between PCPs' attitudes about unhealthy AOD use and their readiness to implement AOD-related preventive care. Our key findings suggest that health professionals' attitudes appear to influence clinical practices, at least self-reported practices, with more stigmatising attitudes associated with lower readiness to implement unhealthy AOD-related preventive care. Moreover, PCPs who endorsed more negative attitudes were also the ones that attributed more blame (responsibility for causing the problem) and control (for the solutions of the problem) to patients with unhealthy AOD use. These findings are consistent with Weiner's attributional theoretical framework [6] and confirm other findings in the literature [3,13].

Health professionals' satisfaction when working with people with unhealthy AOD use was associated with their readiness to implement AOD-related preventive care. In the same direction, Saitz and colleagues [11] showed that health professionals' greater satisfaction caring for patients with substance problems was associated with positive attitudes towards addiction treatment.

Previous research has found that health professionals tend to be positive about working with people with unhealthy AOD use [12,15]; however, this seems to be more associated with a sense of professional responsibility, rather than personal interest, as at the same

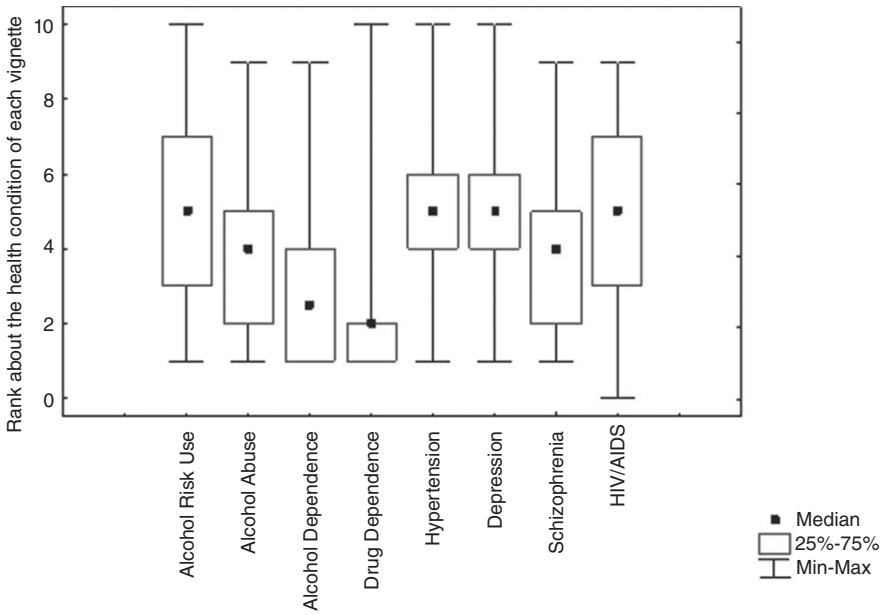


Figure 1. *Primary care professionals rating of the health condition presented in each vignette [from very bad (0) very good (10) health]. *Alcohol abuse rating differs significantly from all other ratings (*$P < 0.01$*) except Schizophrenia. **Alcohol and drug dependence ratings differ significantly (*$P < 0.05$*) from all others.*

time, they tend to report low levels of professional satisfaction when working with this population. This discrepancy between perceived role responsibility and other attitudes and satisfaction seems to be even more evident for professionals' views regarding people with unhealthy drug use in particular, as this study also showed. The illegal status of these substances can reinforce stereotypes of dangerousness, creating safety concerns among health professionals [22–24], leading to discrimination against the patient, and poor quality of care [25].

The present study has some limitations. Causality cannot be determined from a cross-sectional survey and we did not adjust for potential confounders in part because of the relatively small sample size. We assessed readiness to implement AOD-related preventive care, not actual implementation of practices, and similarly, instead of observing actual practice with patients, we used vignettes of hypothetical cases that may or may not reflect how health professionals would feel or respond in real situations. Social desirability and selection bias may also have biased PCPs' responses towards more positive attitudes. For example, we found no association between PCP AOD use and readiness while others have suggest such a relation [26]. On the other hand, the fact that PCPs did report substantial lack of readiness and negative attitudes argues against such biases. In addition, the fact that the survey was anonymous likely minimised these biases. And the response rate was similar to that seen often in surveys of health professionals [24]. The generalisability of this study may be limited as this is a relatively small sample, from one city in Brazil. However, at a minimum the results likely apply to the many health centres in Sao Paulo and other urban centres in Brazil. Also, as the findings seem consistent with studies carried out in other countries [9,12,15], there is little reason to suspect that the findings would differ if the study were repeated elsewhere.

Despite the study limitations, the findings have important practical implications. We found associations between attitudes, including stigmatising attitudes, and clinicians' readiness to implement AOD-related preventive care. We have no reason to believe that these associations would not apply to PCPs in various contexts, and in developing and developed nations alike. These associations suggest that the way unhealthy AOD use is perceived by PCPs has implications for the relationship between the patient and clinician, and for the quality of care they receive. The development of positive attitudes and practices among health professionals in regard to AOD-related work must involve interventions that go beyond the individual [27,28]. Training and education are strategies that may be necessary but not sufficient to change health professionals' attitudes and practices [10,13]. Yet knowing that attitudes are important in determining practices suggests avenues for change and improvement of care. Multifaceted approaches, including education, training, practice guidelines, systems

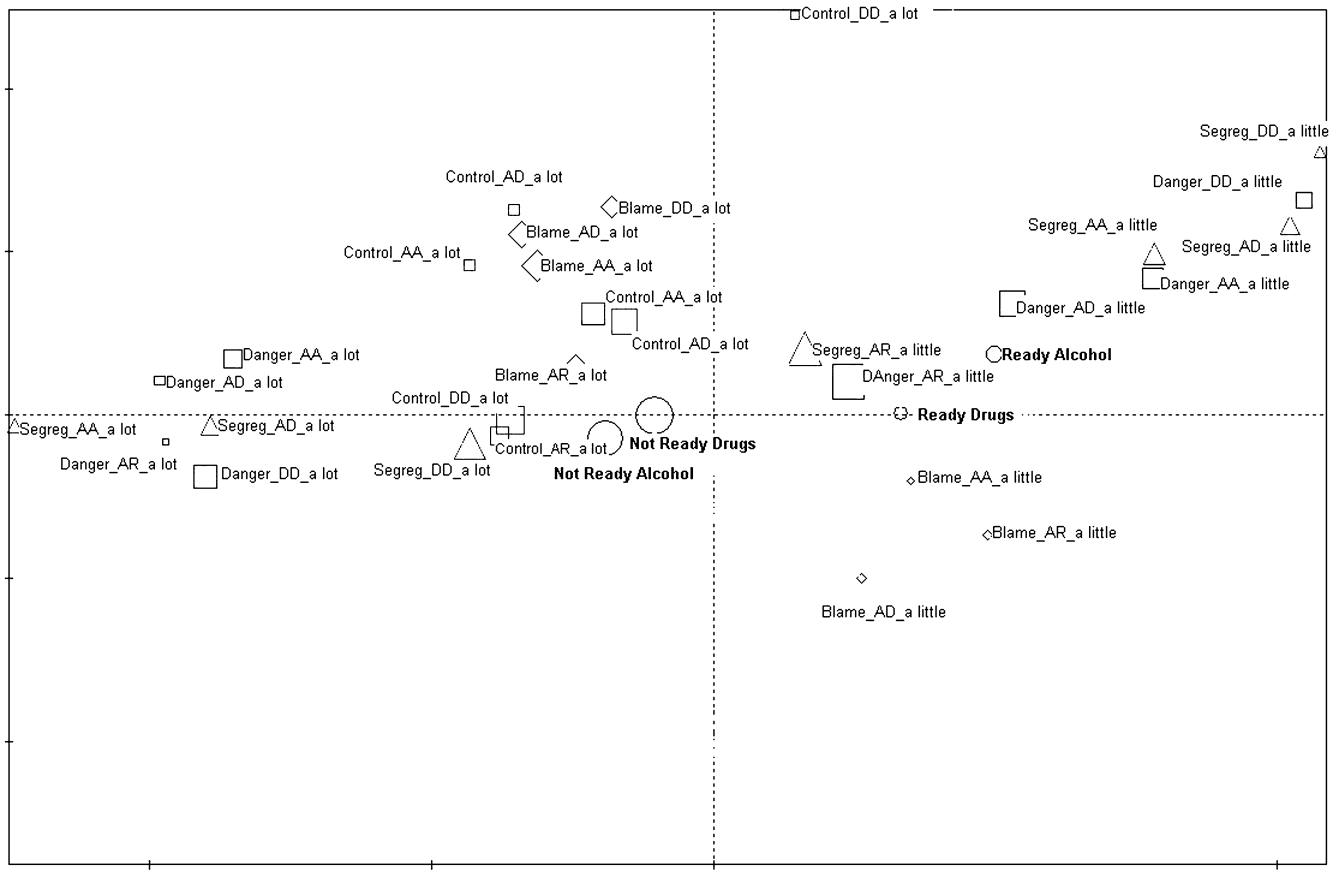


Figure 2. *Multiple Correspondence Analysis map of the associations between the items in the Adapted Attribution Questionnaire and primary care professionals' readiness to implement alcohol and other drug clinical prevention practices. Different geometric figures represent the different variables in the model that are significantly associated with readiness, and the proximities reflect the strength of associations. Vignette conditions: AA, alcohol abuse; AD, alcohol dependence; AR, risky alcohol use; DD, drug dependence. Not ready/Ready, to carry out preventive interventions for risky alcohol/drug use. Control, danger(ousness), segreg(ation) and blame are items from the Adapted Attribution Questionnaire addressing stigmatising attitudes.*

approaches, policy changes and attention to societal norms, may improve PCP's attitudes and have substantial impact on patient care. Such approached will likely need to be adapted to local conditions and circumstances. Efforts in these areas should be pursued and studied to improve the care of primary care patients with unhealthy AOD use.

Acknowledgements

We thank Dr Fernando Colugnati for the statistical analysis and manuscript review. Michaela Bitarello do Amaral-Sabadini (MB Amaral-Sabadini) had financial support from: FAPESP—Fundação de Amparo a Pesquisa do Estado de São Paulo—Brazil (Doctoral Scholarship—Process number: #2007/07368-2) and from CAPES—Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brazil (Scholarship—Process number: # 0389099).

References

[1] Heatherton TF, Kleck RE, Hebl MR, Hull JG. The social psychology of stigma. New York: The Guilford Press, 2003.
[2] Room R. Stigma, social inequality and alcohol and drug use. Drug Alcohol Rev 2005; 24:143–55.
[3] Corrigan PW, Lurie BD, Goldman HH, Slopen N, Medasani K, Phelan S. How adolescents perceive the stigma of mental illness and alcohol abuse. Psychiatr Serv 2005; 56:544–50.
[4] Gilbert DT, Fiske ST, Lindzey G. The handbook of social psychology, 4th edn. New York: Oxford University Press, 1998.
[5] Link, BG, Phelan, JC. Conceptualizing stigma. Annu Rev Sociol 2001; 27:363–85.
[6] Weiner B. Social motivation, justice, and the moral emotions—an attributional approach. Mahwah: Lawrence Erlbaum Associates, 2006.

[7] Corrigan PW, Markowitz FE, Watson AC, Rowan D, Kubiak MA. An attribution model of public discrimination towards persons with mental illness. J Health Soc Behav 2003; 44:162–79.
[8] Palm J. The nature of and responsibility for alcohol and drugs problems: views among treatment staff. Addict Res Theory 2004; 12:413–31.
[9] McCormick KA, Cochran NE, Back AL, Merrill JO, Williams, EC, Bradley KA. How primary care providers talk to patients about alcohol—a qualitative study. J Gen Intern Med 2006; 21:966–72.
[10] Amaral MB, Ronzani TM, Souza-Formigoni MLO. Process evaluation of the implementation of a screening and brief intervention program for alcohol risk in primary health care: an experience in Brazil. Drug Alcohol Rev 2010; 29:162–8.
[11] Saitz R, Friedman PD, Sullivan LM, *et al*. Professional satisfaction experienced when caring for substance-abusing patients—faculty and resident physicians perspectives. J Gen Intern Med 2002; 17:373–6.
[12] Anderson P, Kaner E, Wutzke S, Wensing M, Grol R, Heather N. Attitudes and management of alcohol problems in general practice: descriptive analysis based on findings of a World Health Organization international collaborative survey. Alcohol Alcohol 2003; 38:597–601.
[13] Ronzani TM, Amaral MB, Souza-Formigoni MLO, Babor TF. Evaluation of a training program to implement alcohol screening, brief intervention and referral to treatment in primary health care in Minas Gerais, Brazil. Nord Stud Alcohol Drugs 2008; 25:529–38.
[14] Ronzani TM, Higgins-Biddle J, Furtado EF. Stigmatization of alcohol and other drug users by primary care providers in Southeast Brazil. Soc Sci Med 2009;69:1080–4.
[15] Aira M, Kauhanen J, Larivaara P, Rautio P. Factors influencing inquiry about patient's alcohol consumption by primary health care physicians: qualitative semi-structured interview study. Fam Pract 2003;20:270–5.
[16] WHO (World Health Organization) & Wonca (World Organization of Family Doctors) Report. Integrating mental health into primary care—a global perspective. ISBN 978 92 4 156368 0, 2008.
[17] Brown SA. Factors and measurement of mental illness stigma: a psychometric examination of the attribution questionnaire. Psychiatr Rehabil J 2008; 32:89–94.
[18] WHO—World Health Organization. The ASSIST project—Alcohol, Smoking and Substance Involvement Screening Test. Available at: http://www.who.int/substance_abuse/activities/assist_portuguese.pdf (accessed May 2008).
[19] WHO ASSIST Working Group. The Alcohol, Smoking and Substance Involvement Screening Test (ASSIST): development, reliability and feasibility. Addiction 2002; 97:1183–94.
[20] Humeniuk R, Ali R, Babor TF, *et al*. Validation of the Alcohol, Smoking And Substance Involvement Screening Test (ASSIST). Addiction 2008; 103:1039–47.
[21] Panagiotakos DB, Pitsavos C. Interpretation of epidemiological data using multiple correspondence analysis and log-linear models. J Data Sci 2004; 2:75–86.
[22] McKeown A, Matheson C, Bond C. A qualitative study of GPs' attitudes to drug misusers and drug misuse services in primary care. Fam Pract 2003; 20:120–5.
[23] Ding L, Landen B, Wilson I, Wong M, Shapiro M, Cleary P. Predictors and consequences of negative physician attitudes toward HIV infected injection drug users. Arch Intern Med 2005; 165:618–23.
[24] Fortney J, Mukherjee S, Curran G, Fortney S, Han X, Booth B. Factors associated with perceived stigma for alcohol use and treatment among at-risk drinkers. J Behav Health Serv Res 2004; 31:418–29.
[25] Yokaichiya CM, Figueredo WS, Schraiber LB. Injecting drug users and antiretroviral therapy: perceptions of pharmacy teams. Rev Saude Publica 2007; 41:14–21.
[26] Kaner EFS, Rapley T, May C. 'Seeing through the glass darkly? A qualitative exploration of GPs' drinking and their alcohol intervention practices.' Fam Pract 2006 23:481–7.
[27] Skinner N, Roche AM, Freeman T, Mckinnon A. Health professionals' attitudes towards AOD-related work: moving the traditional focus from education and training to organizational culture. Drugs Educ Prev Policy 2009; 16:232–49.
[28] Pavin T, Duarte P, Souza-Formigoni ML. Predictors of the implementation of SBI by health professionals trained by SUPERA course twenty months before. Available at: http://www.inebria.net/Du14/html/en/dir1338/doc17833.htm (accessed November 2009).

Race Differences in Cardiac Catheterization: The Role of Social Contextual Variables

Brian J. Ayotte, PhD[1,2] and Nancy R. Kressin, PhD[2,3,4]

[1] Center for Organizational, Leadership, and Management Research, VA Boston Healthcare System, Boston, MA, USA; [2] VA Boston Healthcare System, Boston, MA, USA; [3] Center for Health Quality, Outcomes and Economic Research, Bedford VA Medical Center, Boston, MA, USA; [4]Health/care Disparities Research Program, Section of General Internal Medicine, Boston University School of Medicine, Boston, MA, USA.

BACKGROUND: Race differences in the receipt of invasive cardiac procedures are well-documented but the etiology remains poorly understood.

OBJECTIVE: We examined how social contextual variables were related to race differences in the likelihood of receiving cardiac catheterization in a sample of veterans who were recommended to undergo the procedure by a physician.

DESIGN: Prospective observational cohort study.

PARTICIPANTS: A subsample from a study examining race disparities in cardiac catheterization of 48 Black/African American and 189 White veterans who were recommended by a physician to undergo cardiac catheterization.

MEASURES: We assessed social contextual variables (e.g., knowing somebody who had the procedure, being encouraged by family or friends), clinical variables (e.g., hypertension, maximal medical therapy), and if participants received cardiac catheterization at any point during the study.

KEY RESULTS: Blacks/African Americans were less likely to undergo cardiac catheterization compared to Whites even after controlling for age, education, and clinical variables (OR=0.31; 95% CI, 0.13, 0.75). After controlling for demographic and clinical variables, three social contextual variables were significantly related to increased likelihood of receiving catheterization: knowing someone who had undergone the procedure (OR=3.14; 95% CI, 1.70, 8.74), social support (OR=2.05; 95% CI, 1.17, 2.78), and being encouraged by family to have procedure (OR=1.45; 95% CI, 1.08, 1.90). After adding the social contextual variables, race was no longer significantly related to the likelihood of receiving catheterization, thus suggesting that social context plays an important role in the relationship between race and cardiac catheterization.

CONCLUSIONS: Our results suggest that social contextual factors are related to the likelihood of receiving recommended care. In addition, accounting for these relationships attenuated the observed race disparities between Whites and Blacks/African Americans who were recommended to undergo cardiac catheterization by their physicians.

Received August 17, 2009
Revised January 28, 2010
Accepted February 12, 2010
Published online April 10, 2010

KEY WORDS: race; differences; cardiac; catheterization.
J Gen Intern Med 25(8):814–8
DOI: 10.1007/s11606-010-1324-y

Race differences in the receipt of invasive cardiac procedures have been well-documented,[1,2] but their etiology remains poorly understood. Although the magnitude of these differences may vary according to sample characteristics, data sources, and study methodology, the preponderance of research suggests that Blacks/African Americans are less likely to receive invasive diagnostic or revascularization procedures for coronary artery disease compared to Whites. Although all of the sources of these disparities are not known, it is clear that the mechanisms underlying race differences in cardiac care are complex and multidimensional and include individual patient factors (e.g., clinical characteristics, health beliefs), physician factors (e.g., practice patterns, availability of technology, stereotypes), and structural factors (e.g., access to and reimbursement for care).[1,3] Identifying sources of the disparities in cardiac care remains critical given the persistent differences in cardiovascular disease between Whites and Blacks/African Americans, and their contribution to excess mortality among the latter.[4]

One possible explanation for disparities in cardiac care could be patient treatment preference.[5] Black/African Americans are significantly less likely to prefer surgical or invasive treatment for a number of conditions, including spinal conditions,[6] cancer screening,[7] and invasive cardiovascular procedures.[8] Studies have found that Blacks/African Americans are more likely to refuse recommended invasive cardiac procedures,[9] even within the VA where financial issues are not a concern.[10] Few studies, however, have examined factors associated with refusal or non-receipt of recommended cardiac procedures in the context of race. Qualitative studies examining race differences in cardiac patients' treatment decision making suggest three levels of factors associated with decision making among Blacks and Whites: (a) health care system (e.g., number of doctors involved with care); (b) personal/social (e.g., role of family); (c) and physician (e.g., quality of communication).[3,11] In these studies, Blacks/African Americans reported several additional factors, including discrimination in the health care system, religion and faith in God, and the need for a physician who understands their symptoms and complaints. In addition, mistrust of the health care system in general might contribute to treatment preference and refusal of recommended care.[12]

The relationship between social factors and treatment preferences is also important to understand.[5] Ferguson identified a number of social contextual factors associated with treatment preference, including discussions with family members, knowing somebody who had the procedure, and religious beliefs. Similarly, a study by Whittle and colleagues found that knowing somebody who had undergone an invasive cardiac procedure was related to increased willingness to undergo a hypothetical invasive cardiac procedure if it was recommended by a physician.[13] In addition, religiosity and attendance at religious services are important in treatment decision making[14] and may reflect an important source of social support.[15,16] Focusing on social contextual factors is particularly useful given the benefits of interventions aimed at social contextual variables and cardiac care. For example, an intervention involving peer support increased participants' self-care behaviors associated with the prevention of heart disease.[17]

The purpose of the current study was to expand upon prior findings and examine how social contextual variables were related to the actual receipt of cardiac catheterization among Blacks/African Americans and Whites who were recommended to have cardiac catheterization by their physician, while simultaneously controlling for clinical and sociodemographic factors that are also known to influence procedure use. We hypothesized that social contextual variables would account for at least some of the race differences in the receipt of recommended cardiac catheterization, after adjusting for other known confounders.

METHOD

Study Setting and Sample

The study methodology has been described in detail elsewhere.[1] Briefly, the study was conducted at five large, urban, academically-affiliated Department of Veterans Affairs (VA) Medical Centers with on-site catheterization laboratories (Houston, Pittsburgh, Atlanta, Durham, St. Louis). A prospective cohort of patients likely to have coronary artery disease was established by screening the results of all cardiac nuclear imaging studies performed between August, 1999, and January, 2001. Nuclear imaging study results were considered positive if there was any evidence of reversible cardiac ischemia (evidenced by reversible defects or redistribution). A total of 1,045 (23% African American, 77% White) patients were included in the final overall baseline sample (a 76% response rate); we selected a subsample for the present analysis.

We used a two-step process to identify patients recommended for catheterization. First, we identified patients who responded positively to an item asking if he or she was offered the option of cardiac catheterization by their physician. Second, we checked these patient reports against physicians' reports of referral for cardiac catheterization. Only patients whose self-report and physician report both indicated that the patient was offered the option of catheterization were included in the current analyses. In total, 417 patients reported that they were not given the option of catheterization, 187 were missing data on this item, and 160 patients reported being given the option of catheterization while their physician indicated that they were not referred for cardiac catheterization. An additional 44 patients were excluded due to missing data on one or more of the other study variables. The final sample consisted of 237 patients.

Procedure

Data were collected from two serially administered questionnaires which included non-overlapping content: one completed within four weeks after the patient's nuclear imaging study and one completed after the patient reported having received the nuclear imaging study results. Patients were contacted by the study research assistant in person or by phone.

Measures

Sociodemographic Information. Patients were asked about their age, education and self-reported race (0=White, 1=Black).

Clinical and Treatment Variables. Clinical and treatment variables were collected by trained nurses who abstracted each patient's medical records. Clinical variables included cardiac symptoms, and medical history (including prior myocardial infarction, hypertension, diabetes, congestive heart failure, renal dysfunction, and lung disease). Maximal medical therapy was identified by using the definition used by the American College of Cardiology/American Heart Association guidelines for coronary angiography and the management of patients with chronic stable angina.[18,19] Data regarding receipt of catheterization was also collected from patients' medical records. These variables were coded as 0=No and 1=Yes. As part of the study questionnaire, patients were also queried about their anginal symptoms, resulting in two scores: anginal stability (0–100 with higher scores indicating greater stability) and anginal frequency (0–100 with higher scores indicating less frequency). Participants who reported no angina received scores of 100 on both variables.

Social Contextual Variables. We included several social contextual variables in our model. Social support was assessed using three items: (a) satisfaction with family relationships, (b) satisfaction with frequency of social contact with friends and relatives, and (c) satisfaction with frequency of contact with someone the patient trusts and can confide in.[20] Respondents answered yes or no with scores reflecting the number of items to which the respondent answered "yes." This scale had acceptable internal reliability for a brief scale among Whites (Cronbach α=0.60) and Black/African Americans (α=0.62). We also assessed whether patients knew any family or friends who had a heart catheterization (0=No, 1=Yes), how well they knew the person who gave him or her the nuclear imaging results (1=Not at all to 5=Very well), if it would be hard on their family if they were in the hospital for more than a couple of days (1=Not at all to 5=Extremely so), and if their family encouraged them to have a heart catheterization (1=Not at all to 5=Extremely so). Marital status was assessed by self-report and was coded as 0=Not Married, 1=Married. Finally,

we assessed religiosity using one item asking about how often patients attend religious services (1=Never/almost never, 8= Daily/more often).

Analyses

Logistic regression models were fit to test the association of social contextual variables with the receipt of heart catheterization after accounting for other factors that might also influence the outcome, as well as clustering of patients within site of care. The first step of the model included the demographic variables of age, education, and race. The second step added clinical (prior myocardial infarction, hypertension, diabetes, congestive heart failure, renal dysfunction, anginal frequency, anginal severity, and lung disease) and treatment variables (maximal medical therapy). The third and final step included the social contextual variables of marital status, social support, familiarity with person who gave results, knowing somebody who had catheterization, perceived difficulty for family if patient was hospitalized, being encouraged by family to have catheterization, and religiosity. The predictive power of the models was examined by use of the area under the ROC curve (*C* statistic). Values near 0.50 reflect a model with no apparent accuracy, while a value of 1.0 reflects perfect accuracy. Clustering of patients within site of care was not significantly related to the outcome variable so it is not reported for efficiency of presentation.

RESULTS

Sample characteristics are presented in Table 1. In terms of race differences in clinical/treatment variables, fewer Blacks/African Americans reported prior revascularization (22.9%) compared to Whites (37.9%), but significantly more Blacks/African Americans reported hypertension (91.7% versus 77.0%). There were also race differences in social contextual variables, with significantly fewer Blacks/African Americans reporting being married (47.9% versus 62.2%) or knowing a family member or friend who had catheterization (62.5% versus 78.8%). However, Blacks/African Americans reported higher religiosity scores (M=6.18, SD=1.46) than Whites (M= 4.63, SD=2.13). Finally, among this subsample of patients referred for catheterization, fewer Blacks/African Americans ultimately received catheterization (77.1%) compared to Whites (89.7%).

The logistic regression model revealed a number of associations with the likelihood of receiving catheterization among this subsample of patients (see Table 2). The baseline model (*C*= 0.60) including the demographic variables of age, education, and race indicated that Blacks/African Americans were significantly less likely than Whites to have catheterization (OR= 0.34; 95% CI, 0.15, 0.79). Age and education were not associated with the likelihood of receiving catheterization.

Adding the clinical and treatment variables in the second model (*C*=0.67) resulted in a slightly lower but still significant OR for Blacks/African Americans (OR=0.31; 95% CI, 0.13, 0.75) compared to the first step of the model. However, none of the clinical or treatment variables were related to the likelihood of receiving catheterization.

In the third model (*C*=0.79), three of the social contextual variables were related to the likelihood of receiving catheterization. First, knowing somebody who had catheterization was related to an increased likelihood of receiving catheterization (OR=3.14; 95% CI, 1.13, 8.74). Second, higher levels of social support were related to increased likelihood of catheterization (OR=2.05; 95% CI, 1.17, 3.60). Third, higher scores on the item assessing if family members encouraged the patient to have catheterization were related to increased likelihood of receiving

Table 1. Sample Characteristics (N=237)

	Black/African Americans N=48	Whites N=189	p-value
Sociodemographic variables			
Age (mean and SD)	61.58 (10.67)	62.59 (9.55)	0.40
Education (mean and SD)	11.97 (2.68)	12.04 (2.61)	0.81
Clinical variables			
Prior revascularization (% yes)	18.6	37.0	0.01
Prior MI (% yes)	23.5	34.7	0.06
Hypertension (%yes)	89.5	78.4	0.02
Angina (% yes)	70.6	72.8	0.69
Congestive heart failure (% yes)	17.4	16.3	0.87
Diabetes (% yes)	30.2	32.1	0.79
Lung disease (% yes)	20.0	24.9	0.39
Renal dysfunction (% yes)	15.1	10.2	0.25
Maximal medical therapy (% yes)	38.4	43.3	0.46
SAQ Anginal frequency (mean and SD)	60.72 (24.69)	60.60 (23.48)	0.99
SAQ Anginal stability (mean and SD)	48.33 (27.95)	51.08 (26.72)	0.48
Social contextual variables			
Married (% yes)	50.0	63.3	0.03
Family/friend had catheterization (% yes)	61.6	80.5	0.01
Social support (mean and SD)	2.22 (0.87)	2.42 (0.85)	0.06
Procedure hard on family (mean and SD)	1.83 (1.27)	1.78 (1.23)	0.78
Familiar with person who gave results (mean and SD)	2.50 (1.51)	2.30 (1.41)	0.25
Family encouraged procedure (mean and SD)	1.93 (1.32)	2.32 (1.51)	0.03
Attend religious services (mean and SD)	6.08 (2.19)	4.16 (2.66)	0.01
Outcome variable			
Receipt of catheterization (% yes)	66.3	81.5	0.03

Table 2. Factors Associated with the Receipt of Cardiac Catheterization Among Patients Referred for the Procedure (N=237)

Variable	Model		
	1	2	3
Sociodemographic			
Age	1.0 (0.96, 1.03)	1.00 (0.97, 1.04)	1.00 (0.96, 1.04)
Education	0.95 (0.83, 1.07)	0.94 (0.82, 1.08)	0.93 (0.81, 1.06)
Race	0.34 (0.17, 0.67)*	0.29 (0.14, 0.60)*	0.44 (0.19, 1.17)
Clinical			
Prior revascularization		1.13 (0.52, 2.48)	0.94 (0.38, 2.93)
Prior myocardial infarction		1.02 (0.48, 2.18)	0.91 (0.39, 2.12)
Hypertension		1.57 (0.68, 3.62)	1.87 (0.72, 4.83)
Angina		0.78 (0.30, 1.99)	0.55 (0.19, 1.58)
Chronic heart failure		0.96 (0.36, 2.54)	0.93 (0.33, 2.64)
Diabetes		0.57 (0.29, 1.14)	0.55 (0.26, 1.18)
Lung disease		0.52 (0.25, 1.15)	0.44 (0.19, 1.03)
Renal dysfunction		0.44 (0.16, 1.23)	0.39 (0.14, 1.17)
Maximal medical therapy		0.80 (0.40, 1.62)	1.13 (0.51, 2.48)
Anginal frequency		0.99 (0.97, 1.02)	1.00 (0.98, 1.02)
Anginal stability		1.00 (0.99, 1.02)	1.00 (0.99, 1.01)
Social contextual variables			
Married			0.84 (0.38, 1.85)
Family/friend had catheterization			3.83 (1.70, 8.65)*
Social support			2.77 (1.13, 2.80)*
Procedure hard on family			0.98 (0.74, 1.30)
Familiar with person who gave results			0.88 (0.67, 1.12)
Family encouraged procedure			1.43 (1.08, 1.90)*
Religiosity			0.92 (0.80, 1.06)
Area under the ROC curve	0.60	0.67	0.79

Note. *p<0.05.

catheterization (OR=1.45; 95% CI, 1.02, 2.06). Interestingly, the addition of social contextual variables into the model resulted in the race variable becoming non-significant (OR= 0.46; 95% CI, 0.17, 1.29).

DISCUSSION

The purpose of this study was to examine the role of race and social contextual variables on cardiac catheterization use among patients who were recommended to receive the procedure. Results from the models that adjusted for demographic and clinical/treatment variables indicated that Blacks/African Americans were less likely to receive catheterization than Whites. However, this relationship became non-significant when social contextual variables were added to the model, suggesting that social contextual variables might play an important part in the relationship between race and the receipt of recommended cardiac catheterization, among those referred for the procedure.

Our findings are consistent with the notion that social contextual variables can influence the acceptability of medical treatments.[21] Three of the social contextual variables were significantly related to the likelihood of receiving catheterization. First, knowing somebody who had cardiac catheterization was related to increased likelihood of receiving the recommended treatment. This is consistent with previous qualitative findings that knowing somebody who had a procedure influenced cardiac patients' treatment decision making.[3] Second, having family members who encouraged the patient to undergo cardiac catheterization was related to an increase in the likelihood of actually undergoing the procedure. Again, this finding supports previous findings indicating that family members play an important role in cardiac treatment decisions.[3,22] Finally, overall social support was related to increased likelihood of receiving catheterization. This finding is similar to previous research that suggests that social support is related to less aversion to invasive procedures among Blacks/African Americans and Whites.[21]

Interestingly, none of the clinical variables included in our model were related to the receipt of cardiac catheterization. This finding may, however, be due to the homogeneous nature of the sub-sample we examined. All of the patients in our sample were recommended to have cardiac catheterization by their physician. Thus, it is likely that the clinical characteristics of the patients were already considered by the physician prior to recommending catheterization, thus resulting in less variability in clinical variables compared to the parent sample.

The reported results should be interpreted in the context of the study limitations. First, our sample consisted of male veterans using VA care, and the results may not generalize to the general population. Second, although the items were from established measures or were constructed based on findings from focus groups,[23] a number of the social contextual variables were single items, which may limit their reliability and validity. Third, the reasons for the patient not receiving the recommended procedure are unknown. Fourth, the study was conducted within the equal-access VA health care system, thus perhaps limiting the generalizability of the findings to the broader health care system where cost is a greater consideration. Fifth, our sample did not include other minority groups (e.g., Hispanic/Latinos) due to the focus of previous literature on differences between Whites and Black/African Americans and the lack of resources to recruit a sufficient number of Hispanic/Latino patients, thus limiting the study generalizability. Finally, we did not include physician characteristics in our model because we chose to focus primarily on individual-level factors that are related to the receipt of recommended care, and

we also had only limited data on physician characteristics. Factors such as communication skills, years of practice, and patient volume (which we did not assess) might influence patients' decisions to undergo certain procedures.

Despite these limitations, the results of this study have implications for the care of patients who are recommended to receive cardiac catheterization. First, our results indicate that physicians should consider the social context in which patients make treatment decisions, and when facing resistance from patients to adhering to recommendations, to explore the possible negative influences of the patient's social system. For example, physicians could ask patients if family and friends are encouraging a particular decision or behavior or if they know anybody who had a similar procedure. Second, the findings suggest that race differences in the likelihood of undergoing a recommended procedure may be, at least in part, due to social contextual variables, suggesting that clinical treatment decision making processes might capitalize on social contextual variables to encourage patients to have needed procedures. For example, encouraging and assisting family members to talk to patients about the benefits of cardiac catheterizations could increase the chances that the patient follows physician recommendations for such procedures. For example, group interventions involving the patient and his or her family could focus on the benefits of a particular procedure and effective communication techniques for discussing medical decisions (e.g., discuss concerns, treatment values). Likewise, the finding that knowing somebody who has undergone cardiac catheterization increased the likelihood of having the procedure suggests that perhaps pairing patients with similar others who have had the procedure (e.g. peer health educators) could maximize the likelihood of the patient feeling more comfortable with the procedure and thus deciding to have the procedure him or herself. This might be particularly important among Black/African Americans because, according to our data, they are less likely to have family or friends who have undergone the procedure.

CONCLUSION

Despite numerous efforts to understand race differences in cardiac care, they remain persistent and widespread. Identifying correlates of receiving recommended care could reduce these disparities and maximize the likelihood that patients undergo recommended cardiac procedures. Our results suggest that social contextual factors are related to the likelihood of receiving recommended care and that interventions aimed at this dimension may help to reduce race differences in the receipt of recommended care. Future research is needed to verify these findings and to identify specific modifiable targets for interventions aimed at increasing the likelihood of patients receiving recommended cardiac care.

Acknowledgements: *This research was supported by funding from the VA Health Services Research and Development Service (HSR&D; ECV97-022, N. Kressin, P.I.). Dr. Kressin is a Research Career Scientist, VA HSR&D (RCS 02-066-1); Dr. Ayotte is a VA HSR&D Post-doctoral fellow. The views expressed in this article are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs.*

Conflict of Interest: *None disclosed.*

Corresponding Author: *Brian J. Ayotte, PhD; VA Boston Healthcare System, 150 S. Huntington Ave. (152m), Boston, MA 02130, USA (e-mail: brian.ayotte@va.gov).*

REFERENCES

1. **Kressin NR, Chang BH, Whittle J, et al.** Racial differences in cardiac catheterization as a function of patients' beliefs. Am J Public Health. 2004;94:2091–7.
2. **Davis AM, Vinci LM, Okwuosa TM, et al.** Cardiovascular health disparities: a systematic review of the health care interventions. Med Care Res Rev. 2008;64(5 Suppl):29S–100S.
3. **Ferguson JA, Weinberger M, Westmoreland GR, et al.** Racial disparity in cardiac decision making: Results from patient focus groups. Arch Intern Med. 1998;158:1450–3.
4. **Wong MD, Shapiro MF, Boscardin WJ, Ettner SL.** Contribution of major diseases to disparities in mortality. N Engl J Med. 2002;347:1585–92.
5. Institute of Medicine. Assessing potential sources of racial and ethnic disparities in care: Patient-and system-level factors. In: Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care. Washington: National Academy Press; 2003:125–59.
6. **Maynard C, Fisher LD, Passamani ER, Pullum T.** Blacks in the coronary artery surgery study (CASS): race and clinical decision making. Am J Public Health. 1986;76:1446–8.
7. **Arega A, Birkmeyer NJO, Lurie JDN, et al.** Racial variation in treatment preferences and willingness to randomize in the spine patient outcome research trial (SPORT). Spine. 2006;31:2263–9.
8. **Rajapaksa R, Macari M, Bini E.** Racial/ethnic differences in patient experiences with and preferences for computed tomography colonography and optical colonoscopy. Clin Gastroenterol Hepatol. 2007;5:1306–12.
9. **Howard GS, Paterniti DA, Wray NP.** Race and patient refusal of invasive cardiac procedures. JGIM. 2004;19:962–6.
10. **Sedlis SP, Fisher VJ, Tice D, Esposito R, Madmon L, Steinberg EH.** Racial differences in performance of invasive cardiac procedures in a Department of Veterans Affairs Medical Center. J Clin Epidemiol. 1997;50:899–901.
11. **Collins TC, Clark JA, Petersen LA, Kressin NR**. Racial differences in how patients perceive physician communication regarding cardiac testing. Med Care. 2002;40:I-27-I-34.
12. **Klonoff EA.** Disparities in the provision of medical care: an outcome in search of an explanation. J Behav Med. 2009;32:48–63.
13. **Whittle J, Conigliaro J, Good CB, Joswiak M.** Do patient preferences contribute to racial differences in cardiovascular procedure use? J Gen Intern Med. 1997;12:267–73.
14. **Silvestri GA, Knittig S, Zoller JS, Nietert PJ.** Importance of faith on medical decisions regarding cancer care. J Clin Oncol. 2003;21:1379–82.
15. **Krause N.** Church-based social support and health in old age. J Gerontol B Psychol Sci Soc Sci. 2002;57:S332–47.
16. **Krause N.** Church-based social support and mortality. J Gerontol B Psychol Sci Soc Sci. 2006;61:S140–6.
17. **Rose MA.** Evaluation of a peer-education program on heart disease prevention with older adults. Public Health Nurs. 1992;9:242–7.
18. **Gibbons R, Chatterjee K, Daley J, et al.** ACC/AHA/ACP-ASIM guidelines for the management of patients with chronic stable angina: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee on the Management of Patients With Chronic Stable Angina). J Am Coll Cardiol. 1999;33:735–1097.
19. **Scanlon PJ, Faxon DP, Audet AM, et al.** ACC/AHA guidelines for coronary angiography: a report of the American College of Cardiology/American Heart Association task force on practice guidelines (Committee on Coronary Angiography). J Am Coll Cardiol. 1999;33(6):1756–824.
20. **Williams RB, Barefoot JC, Califf RM, et al.** Prognostic importance of social and economic resources among medically treated patients with angiographically documented coronary artery disease. JAMA. 1992;267:520–4.
21. **Bosworth HB, Stechuchak KM, Grambow SC, Oddone EZ.** Patient risk perceptions for carotid endarterectomy: which patients are strongly averse to surgery? J Vasc Surg. 2004;40:86–91.
22. **Whittle J, Conigliaro J, Good CB, Joswiak M.** Do patient preferences contribute to racial differences in cardiovascular procedure use? J Gen Inten Med. 1997;12:267–73.
23. **Kressin NR, Clark JA, Whittle J, et al.** Racial differences in health-related beliefs, attitudes, and experiences of VA cardiac patients. Med Care. 2002;40:I-72-I-80.

Original Article

Predictors of Timely Follow-Up After Abnormal Cancer Screening Among Women Seeking Care at Urban Community Health Centers

Tracy A. Battaglia, MD, MPH[1]; M. Christina Santana, MPH[1]; Sharon Bak, MPH[1]; Manjusha Gokhale, MA[2]; Timothy L. Lash, DSc[3]; Arlene S. Ash, PhD[4]; Richard Kalish, MD, MPH[5]; Stephen Tringale, MD[6]; James O. Taylor, MD[7]; and Karen M. Freund, MD, MPH[1]

BACKGROUND: We sought to measure time and identify predictors of timely follow-up among a cohort of racially/ethnically diverse inner city women with breast and cervical cancer screening abnormalities. **METHODS:** Eligible women had an abnormality detected on a mammogram or Papanicolaou (Pap) test between January 2004 and December 2005 in 1 of 6 community health centers in Boston, Massachusetts. Retrospective chart review allowed us to measure time to diagnostic resolution. We used Cox proportional hazards models to develop predictive models for timely resolution (defined as definitive diagnostic services completed within 180 days from index abnormality). **RESULTS:** Among 523 women with mammography abnormalities and 474 women with Pap test abnormalities, >90% achieved diagnostic resolution within 12 months. Median time to resolution was longer for Pap test than for mammography abnormalities (85 vs 27 days). Site of care, rather than any sociodemographic characteristic of individuals, including race/ethnicity, was the only significant predictor of timely follow-up for both mammogram and Pap test abnormalities. **CONCLUSIONS:** Site-specific community-based interventions may be the most effective interventions to reduce cancer health disparities when addressing the needs of underserved populations. ***Cancer*** **2010;116:913–21.** © *2010 American Cancer Society.*

KEYWORDS: public health, women's health, neoplasms, health services, internal medicine.

Despite increasing gains in cancer care,[1] disparities in cancer outcomes are well documented for racial/ethnic minorities and those of low socioeconomic status.[2] In Massachusetts, non-Hispanic black women have higher mortality from breast cancer than their white counterparts (35.5 and 23.3 per 100,000, respectively).[3] The age-adjusted cervical cancer incidence rates in Massachusetts are 5.8 per 100,000 for white women, 9.2 for black women, and 13.1 for Hispanic women.[4] Differences in access to care along the entire cancer care continuum, from screening through diagnostic care to treatment and survivorship, and barriers to using otherwise accessible care, may contribute to these disparities.

Parity in receipt of cancer screening has been achieved in many settings, including Massachusetts.[5] The Centers for Disease Control-funded National Breast and Cervical Cancer Early Detection Program provides millions of dollars annually to ensure that those most at risk have access to breast and cervical cancer screening.[6] However, the prevention potential of cancer screening requires timely diagnostic follow-up once an abnormality has been detected. Delays in diagnosis and treatment as little as 3 months have been shown to increase recurrence[7] and reduce survival rates.[8,9] The belief that these delays contribute to cancer disparities is evident in the emergence of innovative programs that aim to reduce delays in receipt of cancer care services. In 2005, the National Cancer Institute's Center to Reduce Cancer Health Disparities and the American Cancer Society funded 9 programs to participate in a Cooperative Group (the Patient Navigation Research

Corresponding author: Tracy A. Battaglia, MD, MPH, Assistant Professor of Medicine and Epidemiology, Boston University Schools of Medicine and Public Health, Women's Health Unit, 801 Massachusetts Avenue 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]Data Coordinating Center, Boston University School of Public Health, Boston, Massachusetts; [3]Department of Epidemiology, Boston University School of Public Health, Boston, Massachusetts; [4]Healthcare Research Unit and Section of General Internal Medicine, Boston University School of Medicine, Boston, Massachusetts; [5]Department of Family Medicine, Boston University School of Medicine and South Boston Community Health Center, Boston, Massachusetts; [6]Codman Square Health Center, Boston, Massachusetts; [7]East Boston Neighborhood Health Center, Boston, Massachusetts

We thank Sarah E. Lane for her valuable contributions in the preparation of the manuscript, and the entire Boston Patient Navigation Research Program Research Team and Community Advisory Panel, without whom this study would not have been possible.

DOI: 10.1002/cncr.24851, **Received:** January 16, 2009; **Revised:** June 11, 2009; **Accepted:** July 7, 2009, **Published online** January 5, 2010 in Wiley InterScience (www.interscience.wiley.com)

Program)[10-12] to evaluate patient navigation interventions to reduce time to diagnosis and treatment for at-risk underserved populations with abnormal cancer screening or newly diagnosed cancer.

The time it takes to complete diagnostic evaluation varies widely, with the uninsured or underinsured and racial/ethnic minorities often having the longest delays.[13-18] Relevant literature is limited by a lack of consistency in reported outcomes. Most studies are small, limited to a single site of care, and include diverse socioeconomic strata. Therefore, we sought to describe delays in receipt of diagnostic services for abnormal mammography and Papanicolaou (Pap) test screening among inner city women seeking care at 6 community health centers in Boston, which serves as the baseline cohort for the Boston Patient Navigation Research Program. These centers serve a high proportion of the city's racial and ethnic minority populations and those of lower socioeconomic status.

MATERIALS AND METHODS

Study Design

This study was conducted to provide baseline estimates of time to diagnostic resolution for a prospective multisite intervention study, the Boston Patient Navigation Research Program. One of 9 groups in the national Patient Navigation Research Program Cooperative Group,[10-12] the Boston Patient Navigation Research Program partnered with 6 independent community health centers (CHCs) in Boston to perform the study. Concurrent baseline data were collected via retrospective medical chart review to determine time to diagnostic resolution for women with mammogram and Pap test abnormalities. The Boston University Medical Center Institutional Review Board reviewed and approved this study.

Study Population

Eligible subjects for the baseline Patient Navigation Research Program cohort included adult women with an abnormality detected by a screening Pap test or mammogram performed between January 1, 2004 and December 30, 2005 at 1 of the 6 CHCs. Women were excluded if they were pregnant or younger than 18 years of age at the time of their abnormality. Eligible mammography results included any Breast Imaging Reporting and Data System (BIRADS) score indicating need for follow-up (BIRADS 0, 3, 4, and 5). Eligible Pap test results included any cellular abnormality indicating need for follow-up. These include atypical squamous cells of undetermined significance positive for human papillomavirus (ASCUS/HPV+), low-grade squamous intraepithelial lesion (LGSIL), high-grade squamous intraepithelial lesion (HGSIL), and carcinoma. All subjects with high-grade abnormalities were included (BIRADS 4, 5; HGSIL), and a random sample of low-grade abnormalities (BIRADS 0, 3; ASCUS/HPV+; LGSIL) were used to reach a sample of approximately 100 screened-positive women per site. At sites with fewer than 100 eligible cases, all eligible subjects were included.

Data Collection

Chart abstraction began in July 2006. If an abnormality had not reached diagnostic resolution by the time of abstraction, the patient's chart was reviewed again, if necessary, at least 365 days after the index event to ascertain resolution. The majority of the abstraction was completed using the electronic medical record; occasionally, missing clinical data needed to be abstracted from the paper chart (approximately 4%). One of 2 authors (M.C.S., S.B.) then reviewed data abstraction forms for completeness, accuracy, and internal consistency, and then entered the data into a secure, password-protected study database.

Study Variables

Variables were selected based on both 1) availability in the medical record and 2) consistency with the data dictionary developed by the Design and Analysis Committee of the National Cancer Institute Patient Navigation Research Program, to enable future comparisons with data from other Patient Navigation Research Programs.[12]

Independent Variables

Race/Ethnicity was documented in the electronic medical record as mutually exclusive response values: White, Black/African American, Asian, Native Hawaiian/ Pacific Islander, Native American/ Alaskan Native, Hispanic Latino, or Other. With the exception of White, Black, and Hispanic, the remaining race categories were collapsed into Other because there were too few subjects in the individual racial categories to yield meaningful analyses. For individuals in >1 category, only the first of Hispanic, Black, White, or Other (in that order) was used. Age was calculated from month and year of birth to the date of the screening test. Different age categories were used for the 2 screening populations; each screening population was separately categorized into 1 of 4 clinically relevant age groups. Primary Language was categorized as English, Spanish, or Other. Primary care status was determined by the presence of a named physician appearing in

the electronic medical record. Primary and secondary insurance as documented in the electronic medical record were used to create the following 3 mutually exclusive categories: No Health Insurance, Publicly Financed Health Insurance Only (Medicare and/or Medicaid as sole insurers), or Some Form of Private Health Insurance.

Outcome Variables

Our primary outcome of interest was time from index screening abnormality to diagnostic resolution. *Diagnostic resolution* was defined as definitive tissue diagnosis (biopsy with pathology report) or clinical evaluation (such as colposcopy) indicating no further need for evaluation, in concordance with the Patient Navigation Research Program.[12] Clinical evaluation was included to account for variation in clinical practices. Because of the variability in the number of days to resolution, and the likelihood of outliers that would result in skewed data, we censored this outcome at a maximum of 180 days for outcomes analyses. Because there are clinical implications to delays as short as 90 days, the authors felt that a 180-day cutoff for timely resolution has adequate clinical significance.[19,20] Subjects were categorized as having timely resolution if their diagnostic resolution occurred within 180 days from the index abnormality. For subjects eligible because of a BIRADS 3 result, the earliest date for resolution or "time 0" began 6 months from the date of the index abnormality, because clinical practice guidelines for follow-up call for a repeat mammogram 6 months after the index abnormality.[21]

Data Analysis

Subjects with abnormal mammogram and Pap tests were analyzed separately. In addition to having 2 different clinical screening programs, the 2 study populations differed markedly by age, racial/ethnic distribution, and proportion with a final diagnosis of cancer. Results are presented in parallel here to provide the opportunity to see whether particular CHCs are consistently better (or worse) than others on both types of cancer screening follow-up.

Descriptive statistics were performed to report the sociodemographic characteristics of the 2 study populations, and to determine median time to resolution and the rate of resolution at different time cutoffs. We calculated *P* values within CHC sites using analysis of variance for continuous variables and chi-square for categorical variables.

Univariate Cox proportional hazard ratios (HRs) were generated to test the association of each subject characteristic with "timely resolution," such that larger hazards ratios are associated with shorter time to resolution. For our multivariate analysis, we predicted timely resolution using Cox-proportional hazards modeling. In the final models, we included only those categorical variables for which the group had a significant *P* value ($<.05$) under a univariate Cox model. The CHC site with the largest study enrollment was chosen as the referent group in both cohort regression models for ease of comparison. All analyses were conducted using SAS 9.1 (SAS Institute, Cary, NC). A 2-sided *P* value $<.05$ was considered statistically significant for reporting associations. We hypothesized that systems issues were extremely important, that is, CHC site was a key explanatory variable, not a confounder. To see if nesting with CHCs had confounded the relationship between patient factors (which differed substantially across CHCs) and timely resolution, we also performed regressions treating CHC as a hierarchical clustering variable. Because these analyses did not change any of our findings, they are not reported. Thus, we did not conduct analyses within strata as defined by CHC site.

RESULTS

A total of 997 women were included in the study (523 with mammogram and 474 with Pap test abnormality). Tables 1 and 2 display subject characteristics by CHC site for mammogram and Pap test subjects, respectively. The different age distributions for each cohort reflect recommended screening guidelines for that cancer site. For both screening groups, the majority were nonwhite, with 19% to 27% Hispanic, 33% to 34% black, and 11% to 14% other. Less than 1/3 had private health insurance. In each screening group, about 1/3 spoke a language other than English as their primary language; Spanish was the most common non-English language spoken (about 15%). The majority of screening abnormalities were low grade in their suspicion for cancer, including BIRADS 0 (67%) and BIRADS 3 (25%) for mammogram and LGSIL (87%) for Pap test subjects.

Subjects across CHCs differed significantly on all demographic characteristics, reflecting the singular populations specific to the communities they serve. For example, the proportion of black subjects ranged from 3% to 88% across the 6 CHCs. Those sites with largely white populations, as demonstrated by Health Center D, which had 92% white subjects, also had the lowest rate of private health insurance (9%), demonstrating a socioeconomically disadvantaged group. Although the data are not

Table 1. Demographic Characteristics of Subjects With Abnormal Mammograms in the Boston PNRP Baseline Cohort by Study Site

Characteristics	Community Health Center Site, No. (%)							P^a
	All	A	B	C	D	E	F	
Total	523	71 (14)	107 (20)	36 (7)	106 (20)	99 (19)	104 (20)	
Age, y								≤.001
30-40	23 (7)	8 (11)	10 (9)	2 (6)	7 (7)	0 (0)	8 (8)	
41-50	274 (52)	27 (38)	44 (41)	15 (42)	39 (37)	99 (100)	43 (41)	
51-64	161 (31)	25 (35)	40 (37)	14 (39)	33 (31)	0 (0)	40 (38)	
65+	65 (12)	11 (15)	13 (12)	5 (14)	27 (25)	0 (0)	13 (13)	
Race								≤.001
Hispanic	99 (19)	24 (34)	16 (15)	3 (8)	4 (4)	43 (44)	9 (8)	
Black	171 (33)	2 (3)	25 (23)	12 (33)	3 (3)	38 (38)	91 (88)	
White	195 (37)	38 (53)	43 (40)	10 (28)	98 (92)	3 (3)	3 (3)	
Other[b]	58 (11)	7 (10)	23 (22)	11 (31)	1 (1)	15 (15)	1 (1)	
Language								≤.001
Spanish	75 (14)	16 (23)	11 (11)	2 (6)	3 (3)	37 (38)	6 (6)	
English	333 (64)	47 (66)	71 (66)	25 (69)	85 (80)	23 (32)	73 (70)	
Other[c]	115 (22)	8 (11)	25 (23)	9 (25)	18 (17)	30 (30)	25 (24)	
Insurance								≤.001
No insurance	221 (42)	12 (17)	63 (59)	13 (36)	44 (42)	41 (42)	48 (46)	
Public	188 (36)	30 (42)	31 (29)	14 (39)	52 (49)	30 (30)	31 (30)	
Private	114 (22)	29 (41)	13 (12)	9 (25)	10 (9)	28 (28)	25 (24)	
BIRADS score								≤.001
0	352 (67)	42 (59)	91 (85)	24 (67)	77 (72)	67 (68)	51 (49)	
3	130 (25)	16 (23)	15 (14)	8 (22)	23 (22)	21 (21)	47 (45)	
4, 5	41 (8)	13 (18)	1 (1)	4 (11)	6 (6)	11 (11)	6 (6)	

PNRP indicates Patient Navigation Research Program; BIRADS, Breast Imaging Reporting and Data System.
[a]Each *P* value is from a chi-square test for independence between the distribution of the row categorical variable and the community health center site.
[b]Any race other than Hispanic, black, or white. Due to differences in race/ethnicity reporting, numbers for collapsed racial/ethnic categories are unavailable. Assignment is to 1 category only, with each category dominating the category below it.
[c]Any language other than English or Spanish. Includes Albanian, Arabic, French, Greek, Italian, Polish, and Portuguese.

shown here, we know that this CHC serves a largely immigrant, Albanian population.

During the 1-year of follow-up, 20 breast and 4 gynecological cancers were diagnosed from the abnormal screening tests. Most cancers occurred in patients whose index abnormality was high grade, including BIRADS 4 or 5 on mammography (11 breast cancers), and carcinoma on Pap test (all 4 gynecologic cancers). The remaining breast cancers occurred in women with a low-grade index mammogram result, including BIRADS 0 (8 cancers) and BIRADS 3 (1 cancer).

Time to diagnostic resolution by screening abnormality is displayed in Table 3. Overall, median time to resolution was shorter for mammogram abnormalities compared with Pap test abnormalities (median Day 27 vs Day 85, respectively). Ninety-two percent of all mammogram abnormalities achieved diagnostic resolution by 180 days, compared with only 65% of Pap test abnormalities. However, almost all abnormalities were resolved within 12 months (97% of mammogram and 93% of Pap test abnormalities). Time to resolution differed by screening abnormality; subjects with high-grade mammogram lesion (BIRADS 4,5) had the longest median time to resolution (36 days), followed by BIRADS 0 (28 days). This same pattern was not observed for high-grade Pap test abnormalities; HGSIL had the shortest time to resolution (median Day 56) compared with the lower grade Pap test lesions ASCUS/HPV+ and LGSIL (median Day 89 and Day 84, respectively). We found little variability across CHCs in the time to diagnostic resolution for a breast abnormality (data not shown). Median time to resolution ranged from 24 to 30 days across CHCs. Minimal increase in proportion of resolved abnormalities was noted beyond 6 months for abnormal mammography screening; increases in diagnostic resolution rates from 6 to 12 months across CHCs ranged from 0% to 8%. In contrast, resolution of Pap test abnormalities was more variable across CHCs, with a median time to resolution range of 59 to 181 days, with as many as

Table 2. Demographic Characteristics of Subjects With Abnormal Pap Tests in the Boston PNRP Baseline Cohort by Study Site

Characteristics	Community Health Center Site, No. (%)							P^a
	All	A	B	C	D	E	F	
Total	474	86 (18)	50 (11)	49 (10)	92 (19)	98 (21)	99 (21)	
Age, y								.10
18-21	92 (19)	14 (16)	11 (22)	11 (22)	16 (17)	17 (17)	23 (23)	
22-25	151 (32)	25 (29)	21 (42)	13 (27)	35 (38)	32 (33)	25 (25)	
26-35	130 (27)	27 (31)	15 (30)	7 (14)	25 (27)	28 (29)	28 (28)	
36+	101 (21)	20 (23)	3 (6)	18 (37)	16 (17)	21 (21)	23 (23)	
Race								≤.001
Hispanic	129 (27)	40 (46)	8 (16)	2 (4)	9 (10)	60 (61)	10 (10)	
Black	160 (34)	4 (5)	21 (42)	25 (51)	4 (4)	30 (31)	76 (77)	
White	121 (26)	23 (27)	12 (24)	6 (12)	75 (82)	1 (1)	4 (4)	
Other[b]	64 (14)	19 (22)	9 (18)	16 (33)	4 (4)	7 (7)	9 (9)	
Language								≤.001
Spanish	71 (15)	31 (36)	5 (10)	1 (2)	4 (4)	24 (25)	6 (6)	
English	317 (67)	44 (51)	34 (68)	36 (73)	73 (80)	51 (52)	79 (80)	
Other[c]	86 (18)	11 (13)	11 (22)	12 (25)	15 (16)	23 (23)	14 (14)	
Insurance								≤.001
No insurance	210 (44)	6 (7)	29 (58)	22 (45)	36 (39)	71 (73)	46 (47)	
Public	129 (27)	24 (28)	9 (18)	18 (37)	46 (50)	15 (15)	17 (17)	
Private	135 (28)	56 (65)	12 (24)	9 (18)	10 (11)	12 (12)	36 (36)	
Cervical abnormality								.43
Low grade[d]	414 (87)	71 (83)	46 (92)	45 (92)	80 (87)	83 (85)	89 (90)	
High grade[e]	60 (13)	15 (17)	4 (8)	4 (8)	12 (13)	15 (15)	10 (10)	

Pap indicates Papanicolaou.
[a]Each *P* value is from a chi-square test for independence between the distribution of the row categorical variable and the community health center site.
[b]Any race other than Hispanic, black, or white. Due to differences in race/ethnicity reporting, numbers for collapsed racial/ethnic categories are unavailable. Assignment is to 1 category only, with each category dominating the category below it.
[c]Any language other than English or Spanish. Includes Arabic, Irish, Polish, and Portuguese.
[d]Includes Pap test results: atypical squamous cells of undetermined significance/human papillomavirus+ and low-grade squamous intraepithelial lesion.
[e]Includes Pap test results: high-grade squamous intraepithelial lesion, carcinoma.

an additional third of diagnostic resolution completed between 181 and 365 days after the index abnormality. Examination of differences in race showed greater variation by CHC than by racial category (data not shown).

Tables 4 and 5 present the univariate and multivariate findings from the Cox proportional hazard models predicting timely resolution for mammogram and Pap test abnormalities, respectively. Univariate analysis of subjects with abnormal mammograms found that CHC and BIRADS designation were significantly associated with timely resolution, but neither composite variable was statistically significant in the multivariate (Table 4). However, there were still intergroup differences, with CHC C (HR, 1.56; confidence interval [CI], 1.02-2.38) and CHC F (HR, 1.41; CI, 1.02-1.95) more likely to have timely resolution compared with referent group CHC A, and BIRADS 0 abnormalities less likely to have timely resolution (HR, 0.79; CI, 0.64-0.98) in comparison to BIRADS 3 abnormalities.

Univariate analysis of subjects with abnormal Pap tests found CHC, race, insurance status, and language to be significantly associated with timely resolution (Table 5). In the multivariate model, CHC was the only composite variable predicting timely resolution ($P<.001$). Intergroup differences found 3 sites, CHC C (HR, 0.39; CI, 0.24-0.64), CHC E (HR, 0.53; CI, 0.35-0.81), and CHC F (HR, 0.40; CI, 0.26-0.62), significantly less likely to have timely diagnostic resolution compared with referent group CHC A. In addition, although insurance status was not significant as a composite predictor, those with no health insurance (HR, 0.71; CI, 0.51-0.98) were less likely to have timely follow-up compared with those with private insurance.

DISCUSSION

This study describes delays in diagnostic resolution after an abnormal breast or cervical cancer screening test among a representative population of primarily minority, urban

Table 3. Uncensored Time to Resolution by Type of Screening Abnormality in the Boston PNRP Baseline Cohort

Abnormality	No.	Median Days to Resolution (Q1, Q3)	% Resolved at 6 Months	% Resolved at 12 Months
Mammogram		27		
All	523	27 (15, 52)	92	97
BIRADS 0	352	28 (20, 56)	92	97
BIRADS 3	130	11 (2, 37)	89	94
BIRADS 4, 5	41	36 (10, 57)	95	100
Pap test		85		
All	474	82 (45, 174)	65	93
Low grade	414	85 (47, 174)	65	93
High grade	60	56 (34, 175)	65	92

PNRP indicates Patient Navigation Research Program; Q, quartile; BIRADS, Breast Imaging Reporting and Data System.

women from a homogeneous socioeconomic strata most at risk for adverse cancer outcomes. The diversity in race/ethnicity across the 6 CHCs is typical of the heterogeneity of populations who receive healthcare at urban safety net institutions.[22-24] After a full year of follow-up, diagnostic resolution for all cancer screening abnormalities reached over 90%; however, significant delays existed in those screening abnormalities most likely to lead to a breast cancer diagnosis. Site of care, rather than any sociodemographic characteristic of individuals including race/ethnicity, was the only significant predictor of delay in both cancer screening groups.

We found that any racial/ethnic differences in timely diagnostic resolution were explained by differences in site of care, suggesting that observed differences in timely follow-up may be primarily because of systems issues within each CHC rather than differences in the populations served. These findings are in contrast to much of the published literature, which repeatedly report minority race/ethnicity to predict delays in diagnostic care.[8,13,15,16,18,25-27] This inconsistency may be explained by the homogeneity in socioeconomic status of this cohort, as suggested by the low rates of private health insurance, even among white subjects—a factor that often confounds racial comparisons. Comparisons to this literature are thus limited by differences in study populations, such that many published studies include diverse socioeconomic strata with various methods of controlling for socioeconomic status.[8,15,16] One study did identify location of care as an important determinant of timely follow-up[16]; however, another study including exclusively uninsured and underinsured women did not include such analyses.[13]

In our study, CHCs with more timely resolution outcomes for a cancer screening test often had delayed resolution for the other cancer screening test. This reinforces the presence of systematic issues within CHCs such that a CHC may have systems to address a particular screening disease, but lack resources for another screening disease. This difference may reflect resource constraints of CHCs and how they prioritize their population's healthcare needs. Although each of the 6 CHC sites had similar resources, such as on-site screening mammography and colposcopy services, programmatic and staffing differences surely existed yet were not measured. Observed differences may reflect systems put in place to reach patients who are at highest risk for delayed follow-up (eg, systems tailored to enhance follow-up of cervical cancer screening abnormalities). The same CHC may not be equipped to handle the systems issues for an older population of breast screening abnormalities.

An alternative explanation for differential outcomes by cancer screening site may be inherent differences in the clinical care for breast and cervical cancer screening. Women screened for breast cancer are willingly participating with forethought; the woman must come to the radiology facility, usually after having made an appointment. This may even be reinforced by the media attention paid to breast cancer screening. In contrast, cervical cancer screening may happen during another healthcare or family planning visit, and may not have been purposeful. In fact, the literature shows improved adherence with cervical cancer screening in vulnerable populations when the screening is done during urgent care visits,[28] yet finds longer delays in follow-up care after screening is done in these urgent care settings.[29]

Finally, subjects across CHCs may differ on their perceptions of cancer risk, which could affect their timely resolution of cancer screening. Although some studies

Table 4. Predictors of Timely Resolution of Mammography Abnormality[a] in the Boston PNRP Baseline Cohort

Characteristic	Univariate, HR (95% CI)	Multivariate, HR (95% CI)	*P*[b]
Age, y			
30-40	1.05 (0.70, 1.59)		
41-50	0.78 (0.59, 1.02)		
51-64	0.79 (0.58, 1.06)		
65+ (ref)	—		
Race			
Hispanic	1.03 (0.28, 1.33)		
Black	1.11 (0.89, 1.38)		
White (ref)	—		
Other	1.22 (0.90, 1.66)		
Language			
Spanish	0.90 (0.79, 1.17)		
English (ref)	—		
Other	0.86 (0.69, 1.08)		
Insurance			
No insurance	0.88 (0.68, 1.12)		
Public	0.98 (0.80, 1.20)		
Private (ref)	—		
BIRADS			.10
0	**0.79 (0.64, 0.98)[c]**	**0.79 (0.64, 0.98)[c]**	
3 (ref)	—	—	
4, 5	0.86 (0.60, 1.32)	0.94 (0.65, 1.36)	
CHC site			.11
A (ref)	—	—	
B	1.27 (0.92,1.74)	1.34 (0.97, 1.86)	
C	1.50 (0.99,2.28)	**1.56 (1.02, 2.38)[c]**	
D	1.25 (0.91, 1.72)	1.28 (0.93, 1.77)	
E	1.04 (0.75, 1.43)	1.05 (0.76, 1.46)	
F	**1.42 (1.04, 1.96)[c]**	**1.41 (1.02, 1.95)[c]**	

PNRP indicates Patient Navigation Research Program; HR, hazard ratio; CI, confidence interval; ref, reference; BIRADS, Breast Imaging Reporting and Data System; CHC, community health center.
[a] Larger hazards ratios are associated with shorter time to resolution.
[b] *P* value is from a chi-square test for model fit between the row categorical variable and the outcome.
[c] Statistically significant finding.

Table 5. Predictors of Timely Resolution of Pap Test Abnormality[a] in the Boston PNRP Baseline Cohort

Characteristic	Univariate, HR (95% CI)	Multivariate, HR (95% CI)	*P*[b]
Age, y			
18-21	0.79 (0.55, 1.12)		
22-25	0.88 (0.64, 1.20)		
26-35	1.05 (0.77, 1.44)		
36+ (ref)	—		
Race/ethnicity			.35
Hispanic	1.20 (0.90, 1.61)	1.36 (0.87, 2.11)	
Black	**0.62 (0.46, 0.84)[c]**	1.10 (0.74, 1.65)	
White (ref)	—	—	
Other	0.90 (0.63, 1.31)		
Language			.24
Spanish	**1.82 (1.35, 2.45)[c]**	1.35 (0.88, 2.08)	
English (ref)	—	—	
Other	1.18 (0.88, 1.59)	1.22 (0.89, 1.67)	
Insurance			.09
No insurance	0.84 (0.62, 1.12)	**0.71 (0.51, 0.98)[c]**	
Public insurance	**0.71 (0.54, 0.93)[c]**	0.77 (0.58, 1.04)	
Private (ref)	—	—	
Pap test			
Low grade	—		
High grade	1.2 (0.83, 1.63)		
CHC site			<.001
A (ref)	—	—	
B	**0.62 (0.41, 0.93)[c]**	0.70 (0.44, 1.11)	
C	**0.37 (0.24, 0.57)[c]**	**0.39 (0.24, 0.64)[c]**	
D	**0.69 (0.49, 0.96)[c]**	0.76 (0.52, 1.12)[c]	
E	**0.55 (0.39, 0.77)[c]**	**0.53 (0.35, 0.81)[c]**	
F	**0.34 (0.24, 0.50)[c]**	**0.40 (0.26, 0.62)[c]**	

PNRP indicates Patient Navigation Research Program; HR, hazard ratio; CI, confidence interval; ref, reference; CHC, community health center.
[a] Larger hazards ratios are associated with shorter time to resolution.
[b] *P* value is from a chi-square test for model fit between the row categorical variable and the outcome.
[c] Statistically significant finding.

have found differences in the perception of cancer risk in the different ethnic populations,[30] survey data from the same health centers failed to identify differences in cancer risk perception (T. A. Battaglia et al, unpublished data).

We found sociodemographic characteristics of study subjects to differ substantially across CHCs, reflecting the sociocultural differences among populations served at health centers within the same zip code. Even within CHCs, we found differences in sociodemographic characteristics for the 2 cancer screening populations, reflecting generational shifts in the community populations given the changing composition of modern cities and immigration patterns. These findings are particularly important given the emerging role of the community health center model in caring for the country's most vulnerable populations.[22-24,31] Our findings highlight the potential diversity both across and within CHCs, underscoring the need to understand the specific sociodemographics of the populations served.

Well over 90% of our subjects achieved diagnostic resolution, supporting the Institute of Medicine's 2002 recognition of the importance of CHCs in increasing access to care and in improving health outcomes for all patients, especially minorities.[32] Our findings also support the notion that absence of racial disparities may be related to CHCs' culturally sensitive practices and community involvement—features that other primary care settings may lack—and speaks to the success of CHC

models in improving health outcomes for these most vulnerable patients.[23,33,34]

It is important to note that we found the longest delays in follow-up occurred among those most likely to be diagnosed with breast cancer (BIRADS 0, 4, or 5), although this was not true for Pap test abnormalities. This may reflect inherent differences in perception of meaning for different cancer screening abnormalities, including fear of possible cancer. Alternatively, it may reflect system issues in accessing timely breast imaging. Although there is no consensus regarding how long a delay ultimately impacts outcomes, it is clinically feasible that these delays may be a mechanism for the persistent gap in cancer outcomes for vulnerable populations. As such, they speak to the need for community-based interventions targeting such at-risk groups. Patient navigation, an emerging model to address cancer health disparities, is an example of a promising community-based approach to address this gap.[11,12]

This study has several limitations, principally that data were collected by retrospective chart review at a single institution for each subject. Specifically, we were limited by CHC record keeping for the years 2004 through 2006; therefore, subjects who achieved diagnostic resolution outside the system will be misclassified as unresolved. In addition, other site-level measures, such as specific funding resources for site-specific cancer programs, were not available from the medical record review. Demographic information was collected at the time of chart abstraction, whereas the abnormality had occurred earlier; thus, for example, our insurance information may not reflect the status at the time the abnormality occurred. We are unaware of any major changes to healthcare coverage in the state during this time period. Provider-level cluster analyses were not performed, as the study sites were not provider-specific systems.

Conclusions

This study found that delays in diagnostic resolution after an abnormal screening test in an urban safety net system are most strongly associated with site of care. Our data support the need for community-based interventions, such as patient navigation, which are culturally targeted, to close the gap in cancer health disparities.

CONFLICT OF INTEREST DISCLOSURES

This study was funded by the National Cancer Institute (U01 CA116892-01) and the Susan G. Komen Foundation (POP-05-04,003).

REFERENCES

1. Jemal A, Thun MJ, Ries LA, et al. Annual Report to the Nation on the Status of Cancer, 1975-2005, Featuring Trends in Lung Cancer, Tobacco Use, and Tobacco Control. *J Natl Cancer Inst*. 2008;100:1672-1694.
2. American Cancer Society. Cancer Facts & Figures 2008. Atlanta, GA: American Cancer Society; 2008.
3. Massachusetts Department of Public Health, Bureau of Health Information, Statistics, Research and Evaluation. Massachusetts Deaths, 2006. Available at: http://www.mass.gov/Eeohhs2/docs/dph/research_epi/death_report_06.pdf Accessed October 17, 2008.
4. Massachusetts Department of Public Health, Bureau of Health Information, Statistics, Research and Evaluation. Cancer Incidence and Mortality in Massachusetts, 2001-2005. Available at: http://www.mass.gov/Eeohhs2/docs/dph/cancer/registry_statewide_01_05_report.pdf Accessed October 17, 2008.
5. Centers for Disease Control and Prevention. Behavioral Risk Factor Surveillance System Survey Data. Atlanta, GA: US Department of Health and Human Services, Centers for Disease Control and Prevention; 2006. Available at: http://apps.nccd.cdc.gov/brfss/, 2008 Accessed October 17, 2008.
6. Centers for Disease Control and Prevention, Division of Cancer Prevention and Control. National Breast and Cervical Cancer Early Detection Program [website]. June 30, 2008. Available at: http://www.cdc.gov/cancer/nbccedp/. Accessed October 17, 2008.
7. Gold HT, Do HT, Dick AW. Correlates and effect of suboptimal radiotherapy in women with ductal carcinoma in situ or early invasive breast cancer. *Cancer*. 2008;113:3108-3115.
8. 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.
9. Hershman DL, Wang X, McBride R, Jacobson JS, Grann VR, Neugut AI. Delay of adjuvant chemotherapy initiation following breast cancer surgery among elderly women. *Breast Cancer Res Treat*. 2006;99:313-321.
10. National Cancer Institute. Patient Navigation Research Program [website]. Available at: http://crchd.cancer.gov/pnp/pnrp-index.html Accessed October 29, 2008.
11. Wells KJ, Battaglia TA, Dudley DJ, et al. Patient navigation: state of the art or is it science? *Cancer*. 2008;113:1999-2010.
12. Freund KM, Battaglia TA, Calhoun E, et al. The NCI Patient Navigation Research Program: methods, protocol, and measures. *Cancer*. 2008;113:3391-3399.
13. Caplan LS, May DS, Richardson LC. Time to diagnosis and treatment of breast cancer: results from the National Breast and Cervical Cancer Early Detection Program, 1991-1995. *Am J Public Health*. 2000;90:130-134.
14. Benard VB, Lawson HW, Eheman CR, Anderson C, Helsel W. Adherence to guidelines for follow-up of low-grade cytologic abnormalities among medically underserved women. *Obstet Gynecol*. 2005;105:1323-1328.
15. Chang SW, Kerlikowske K, Napoles-Springer A, Posner SF, Sickles EA, Perez-Stable EJ. Racial differences in timeliness of follow-up after abnormal screening mammography. *Cancer*. 1996;78:1395-1402.

16. Press R, Carrasquillo O, Sciacca RR, Giardina EG. Racial/ethnic disparities in time to follow-up after an abnormal mammogram. *J Womens Health (Larchmt)*. 2008;17:923-930.
17. Khanna N, Phillips MD. Adherence to care plan in women with abnormal Papanicolaou smears: a review of barriers and interventions. *J Am Board Fam Pract*. 2001;14:123-130.
18. Engelstad LP, Stewart SL, Nguyen BH, et al. Abnormal Pap smear follow-up in a high-risk population. *Cancer Epidemiol Biomarkers Prev*. 2001;10:1015-1020.
19. 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;23:6639-6646.
20. Hershman DL, Wang X, McBride R, Jacobson JS, Grann VR, Neugut AI. Delay in initiating adjuvant radiotherapy following breast conservation surgery and its impact on survival. *Int J Radiat Oncol Biol Phys*. 2006;65:1353-1360.
21. National Comprehensive Cancer Network. NCCN Practice Guidelines in Oncology-v. 1.2008. Breast Cancer Screening and Diagnosis. 2008. Available at: http://www. nccn.org/professionals/physician_gls/PDF/breast-screening.pdf Accessed October 23, 2008.
22. Politzer RM, Schempf AH, Starfield B, Shi L. The future role of health centers in improving national health. *J Public Health Policy*. 2003;24:296-306.
23. Shi L, Stevens GD, Wulu JT Jr, Politzer RM, Xu J. America's Health Centers: reducing racial and ethnic disparities in perinatal care and birth outcomes. *Health Serv Res.* 2004;39(6 pt 1):1881-1901.
24. Health Centers' Role in Reducing Racial and Ethnic Health Disparities (Fact sheet #0508). Bethesda, MD: National Association of Community Health Centers; 2008.
25. Eggleston KS, Coker AL, Das IP, Cordray ST, Luchok KJ. Understanding barriers for adherence to follow-up care for abnormal Pap tests. *J Womens Health (Larchmt)*. 2007;16:311-330.
26. Fox P, Amsberger P, Zhang X. An examination of differential follow-up rates in cervical cancer screening. *J Community Health*. 1997;22:199-209.
27. Jones BA, Dailey A, Calvocoressi L, et al. Inadequate follow-up of abnormal screening mammograms: findings from the race differences in screening mammography process study (United States). *Cancer Causes Control*. 2005;16:809-821.
28. Batal H, Biggerstaff S, Dunn T, Mehler PS. Cervical cancer screening in the urgent care setting. *J Gen Intern Med*. 2000;15:389-394.
29. Megevand E, Van Wyk W, Knight B, Bloch B. Can cervical cancer be prevented by a see, screen, and treat program? A pilot study. *Am J Obstet Gynecol*. 1996;174:923-928.
30. Kim SE, Perez-Stable EJ, Wong S, et al. Association between cancer risk perception and screening behavior among diverse women. *Arch Intern Med*. 2008;168:728-734.
31. Eisert SL, Mehler PS, Gabow PA. Can America's urban safety net systems be a solution to unequal treatment? *J Urban Health.* 2008;85:766-778.
32. Smedley BD, Stith AY, Nelson AR. Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care. Washington, DC: National Academies Press; 2003.
33. Poon EG, Haas JS, Louise Puopolo A, et al. Communication factors in the follow-up of abnormal mammograms. *J Gen Intern Med*. 2004;19:316-323.
34. Coker AL, Eggleston KS, Meyer TE, Luchok K, Das IP. What predicts adherence to follow-up recommendations for abnormal Pap tests among older women? *Gynecol Oncol.* 2007;105:74-80.

Editorial

The Clock Is Ticking: The Case for Achieving More Rapid Control of Hypertension

Dan R. Berlowitz, MD, MPH;[1,2] Stanley Franklin, MD;[3]

Start low, go slow for many years has been the paradigm guiding drug treatment of hypertension. Patients start on low doses of a single medication; if control is not achieved the dosage is gradually increased or additional medications are started. Yet studies persistently show that by this method the majority of patients with hypertension do not achieve adequate blood pressure control.[1] While there are many reasons, including patient nonadherence with recommended therapies and resistant hypertension,[2,3] a growing literature highlights the role of therapeutic inertia.[4,5] Consistent with the *go slow* paradigm, patients with elevated blood pressures are seen repeatedly by clinicians but therapy is not intensified. As a result, many patients with inadequate blood pressure control are treated with only one or two antihypertensive medications, often at low dosages, for prolonged periods.

The evidence now suggests that to improve cardiovascular outcomes, we require a new paradigm that emphasizes rapid achievement of blood pressure control.[6] We believe that central to this paradigm should be an explicit expectation of the timeframe in which blood pressure control should be achieved. While existing guidelines increasingly emphasize the use of drug combinations to achieve more rapid control, they are surprisingly quiet on this issue of time to control. The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure recommends: "once antihypertensive drug therapy is initiated, most patients should return for follow-up and adjustment of medications at approximately monthly intervals until the blood pressure goal is achieved," and that more frequent visits will be necessary for patients with stage 2 hypertension.[1] Recently published guidelines from the European Society of Hypertension and European Society of Cardiology make no reference to the time between follow-up visits or to achieve control.[7] We now argue for an explicit expectation that anticipates blood pressure control within 3 months of initiating therapy.

From the Center for Health Quality, Outcomes and Economic Research, Bedford VA Hospital, Bedford, MA;[1] Boston University Schools of Public Health and Medicine, Boston, MA;[2] and the Department of Medicine, University of California Irvine, Irvine, CA[3]
Address for correspondence:
Dan R. Berlowitz, MD, MPH, CHQOER (152), Bedford VA Hospital, 200 Springs Road, Bedford, MA 01730
E-mail: dberlow@bu.edu

doi: 10.1111/j.1751-7176.2010.00274.x

BENEFITS OF ACHIEVING RAPID BLOOD PRESSURE CONTROL

Recent hypertension clinical trials with major cardiovascular endpoints have highlighted two key facts about blood pressure control. First, early differences in achieved blood pressures among comparative treatment arms, even when managed by experienced clinician-investigators, are not easily eliminated despite continuing efforts to reach specified targets. Thus, initial failure to treat aggressively makes it less likely that control will ever be achieved. Second, the resulting small differences in achieved blood pressures account for significant differences in clinical event rates.

One strong example of this was the Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT), in which high risk hypertensive patients randomized to diuretic-based treatment had persistently lower blood pressures than those starting with either an angiotensin-converting enzyme (ACE) inhibitor or a calcium channel blocker.[8] The resulting blood pressure inequalities among the treatment groups, which were most prominent in the early study months, remained throughout the 5 year trial despite repeated exhortations to investigators to seek blood pressure control.

The effects on major cardiovascular endpoints of the speed and completeness of blood pressure control were explored in the Valsartan Antihypertensive Long-Term Use Evaluation (VALUE) trial.[9] In a cohort of hypertensive patients at high cardiovascular risk, a calcium channel blocker achieved greater blood pressure reductions in the early part of the study than an angiotensin receptor blocker, which persisted for the 5-year duration of the study. This difference in achieved blood pressure appeared to account for significant differences in clinical events; when patients from the two study groups were matched on the basis of identical achieved blood pressures as well as other key clinical and demographic features, there were no endpoint differences between the treatments.[10] The importance of achieving systolic blood pressure control (<140 mm Hg), regardless of drug assignment, within the study's initial 6-month treatment titration period was also demonstrated. Subsequent fatal and nonfatal cardiovascular events were sharply lower, with a hazard ratio of 0.75 in patients achieving blood pressure control as compared with those who did not.[9] Indeed, there was an indication that blood pressures achieved by just 1 month of study treatment were predictive of outcomes during the following 5 years.

Another high profile hypertension outcomes study affected by early blood pressure differences between treatment groups was the Anglo-Scandinavian Cardiac Outcomes Trial (ASCOT) in which the amlodipine-perindopril arm had a 5.9/2.4 mm Hg lower blood pressure at 3 months compared to atenolol-bendroflumethiazide.[11] This trial was stopped earlier than planned when the safety committee noted that mortality and other endpoints were higher in patients randomized to β-blocker/thiazide treatment than in those randomized to calcium channel blocker/ACE inhibitor treatment. At least part of these differences in event rates could be explained by the differences in blood pressure control.[11]

SAFETY OF RAPID BLOOD PRESSURE CONTROL

A concern when starting antihypertensive therapy is an excessive fall in blood pressure. Certainly, this was an issue with the drugs used 3 or 4 decades ago that worked primarily by interrupting the sympathetic nervous system. The traditional injunction to *start low, go slow* was appropriate with drugs that could produce dizziness and postural hypotension. This concern, however, is less justified with modern drugs. For instance, in the registration studies performed with the combination of irbesartan and hydrochlorothiazide in patients with severe hypertension (baseline diastolic blood pressures >110 mm Hg), there were few symptomatic complaints when therapy was started with this combination.[12] In only 3 out of 328 patients was hypotension observed, and even then only after protocol-mandated up-titration to the maximum dose in patients who had already achieved major blood pressure reductions with the starting dose. A further study in patients with less severe hypertension provided additional reassurance.[13] Despite lower baseline blood pressure values, there was a similarly low incidence of observed hypotension, again only after mandated titration to the maximum combination dose. A recent report of a clinical trial in which hypertensive patients were exposed to initial therapy with maximum doses of a calcium channel blocker/angiotensin receptor blocker combination has provided further confirmation of safety.[14]

Safety concerns may be especially great in elderly populations with isolated systolic hypertension—a group with the lowest rates of blood pressure control and the greatest absolute benefit with effective blood pressure reduction.[1] Although lower initial medication doses may sometimes be indicated to minimize symptoms in elderly hypertensive patients, ultimately most will require and tolerate standard doses and multiple drugs to reach blood pressure targets. In the Hypertension in the Very Elderly Trial (HYVET) (≥80 years), combination therapy was used in almost 3 out of 4 participants because of the difficulty in achieving goal blood pressure even though the target systolic blood pressure was <150 mm Hg, and side effects were no more frequent than in the control group.[15] There are no randomized clinical trials to date supporting lowering systolic blood pressure below 140 mm Hg in elderly patients.[16]

Monotherapy with a diuretic, as recommended in past and present guidelines for initiation of therapy, may in fact create greater safety issues than

starting with a combination. Diuretics can produce hypovolemia, creating a risk of acute hypotension when blockers of the renin-angiotensin system are later added. Thiazides as single agents can also cause unwanted metabolic changes like hypokalemia or hyperglycemia, which are at least partly prevented by concomitant treatment with renin-angiotensin system blockers.[17]

FEASIBILITY OF ACHIEVING BLOOD PRESSURE CONTROL WITHIN 3 MONTHS

Results from early clinical trials may have misleadingly conveyed the impression that blood pressure control cannot be achieved despite years of therapy. In ALLHAT, control was achieved in 66% of patients after 5 years of follow-up.[8] However, control rates were only 50% at 6 months and 55% at 1 year. Moreover, even in the rigorous setting of this clinical trial, failures to intensify therapy despite inadequate control were frequent and it was judged "very likely that better blood pressure control rates could have been achieved if therapy were intensified more consistently for persistent systolic blood pressures $\geq$140 mm Hg."[8]

Higher rates of control in shorter time periods have been seen in more recent clinical trials. For example, a recent review of 9 clinical trials highlighted that among hypertensive patients with diastolic blood pressures between 95 and 110 mm Hg, control (<140/90 mm Hg) was achieved within 8 weeks for 74.6% of patients initiating valsartan 160 mg plus hydrochlorothiazide and in 84.8% starting valsartan 320 mg plus hydrochlorothiazide, without any additional titrations.[18] Among patients with uncontrolled hypertension on monotherapy in the Irbesartan/HCTZ Blood Pressure Reductions in Diverse Patient Populations (INCLUSIVE) trial, use of a combination angiotensin receptor blocker/thiazide diuretic with a single dose titration resulted in overall control rates of 69% in 18 weeks; control rates were considerably higher in the 70% of the study population without diabetes and its stricter definition of blood pressure control.[19] In the Avoiding Cardiovascular Events in Combination Therapy in Patients Living With Systolic Hypertension (ACCOMPLISH) trial, a comparison of the two drug combinations benazepril plus hydrochlorothiazide and benazepril plus amlodipine, with dose increases of the study medications after 1 and 2 months of therapy, control rates in excess of 70% were achieved by 3 months.[14] With additional medication increases after 3 months, control rates in the US cohort in ACCOMPLISH reached 78% by 6 months.

Common to these newer trials is the initiation of therapy with two antihypertensive drugs and mandated dose increases if control is not achieved within several weeks. The need for initial combination therapy is particularly important for grade 2 hypertension and in patients with diabetes and chronic renal disease, with its lower target goal of <130/80 mm Hg.[1] Even with combination therapy, achieving blood pressure control in 3 months or less may be difficult. It will require multiple visits, as often as every 2 weeks, in order to monitor blood pressure and the impact of therapy intensification. This becomes practical and safe as forced titration experience, such as the irbesartan/thiazide combination studies to rapidly control severe and moderate hypertension, have shown that approximately 60% of the blood pressure reduction can be obtained within 2 weeks of initiating therapy.[12,13]

IMPLICATIONS OF AN EXPLICIT EXPECTATION FOR RAPID BLOOD PRESSURE CONTROL

Care for hypertensive patients will need to be changed to meet this expectation. Along with the broader use of combinations, physician practices will need to be organized so as to allow the rescheduling of hypertensive patients on intervals as short as 2 weeks. Home self-measurement of blood pressure will be especially useful for assessing responses to antihypertensive medications, improving patient adherence, and helping to diagnose white-coat effects and masked hypertension[20,21]; the former requires restraint in medicating and the latter perhaps more aggressive therapy than indicated by office blood pressure readings.[22,23] A home blood pressure monitor with a graphic display of weekly control rates was associated in one study with more rapid blood pressure control than a standard monitor.[24]

An explicit expectation for blood pressure control within 3 months could have a number of beneficial effects on hypertensive patients. It would likely lead to more rapid and better blood pressure control with resulting reductions in cardiovascular morbidity and mortality. Achieving blood pressure control earlier may reduce the overall need for frequent office visits later, and hence, reduce overall costs. Such an expectation is easily measurable and could be readily incorporated into a national performance measure. Quality improvement efforts focusing on this measure would promote further improvements in care through the audit and feedback of clinicians' practices. However, there also could be detrimental effects. Patients with

hypertension could be overmedicated, although home blood pressure monitoring might minimize this problem. Inclusion as a quality measure may be associated with unintended consequences such as the diversion of resources away from other important aspects of care.[25] Holding clinicians accountable to a 3-month threshold could also penalize clinicians providing excellent care to very difficult to control patients.[26] Groups developing national quality measures will need to carefully balance the benefits and risks of incorporating this expectation into formal measures.

CONCLUSIONS

We believe that the balance of the evidence supports changing the paradigm of hypertension treatment and implementing an expectation that blood pressure control will be achieved within 3 months of starting medication therapy. Such an explicit expectation would help guide clinicians and provide a benchmark for which they could strive. *Start low, go slow* should no longer serve as the model guiding drug therapy of hypertension. To maximize reductions in cardiovascular events, care for patients with severe or complicated hypertension should, right from the start, anticipate the more forceful and frequent interventions now supported by clinical trial evidence. *Move fast, take control* should now guide clinicians' management of hypertension.

Disclosures: Dr Berlowitz has received past grant support from Bristol-Myers-Squibb and Sanofi-Aventis; he has been a past consultant to Bristol-Myers-Squibb. Dr Franklin is a consultant for AtCor Medical and Bristol-Myers-Squibb.

REFERENCES

1 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. *JAMA*. 2003;289:2560–2572.
2 Garg JP, Elliott WJ, Folker A, et al. Resistant hypertension revisited: a comparison of two university-based cohorts. *Am J Hypertens*. 2005;18:619–626.
3 Rose AJ, Berlowitz DR, Orner MB, et al. Understanding uncontrolled hypertension: is it the patient or provider? *J Clin Hypertens (Greenwich)*. 2007;9:937–943.
4 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.
5 Okonofua EC, Simpson KN, Jesri A, et al. Therapeutic inertia is an impediment to achieving the Healthy People 2010 blood pressure control goals. *Hypertension*. 2006; 47:345–351.
6 Basile J. The importance of prompt blood pressure control. *J Clin Hypertens (Greenwich)*. 2008;10(suppl 1):13–19.
7 Mancia G, De Backer G, Dominiczak A, et al; for the Management of Arterial Hypertension of the European Society of Hypertension; European Society of Cardiology. 2007 Guidelines for the Management of Arterial Hypertension: The Task Force for the Management of Arterial Hypertension of the European Society of Hypertension (ESH) and of the European Society of Cardiology (ESC). *J Hypertens*. 2007;25:1105–1187.
8 Cushman WC, Ford CE, Cutler JA, et al. Success and predictors of blood pressure control in diverse North American settings: the antihypertensive and lipid-lowering treatment to prevent heart attack trial. *J Clin Hypertens (Greenwich)*. 2002;4:393–404.
9 Julius S, Kjeldsen E, Weber MA, et al. Outcomes in hypertensive patients at high cardiovascular risk treatment with regimens based on valsartan or amlodipine: the VALUE randomised trial. *Lancet*. 2004;363:2022–2031.
10 Weber MA, Julius S, Kjeldsen SE, et al. Blood pressure dependent and independent effects of antihypertensive treatment on clinical events in the VALUE Trial. *Lancet*. 2004;363:2049–2051.
11 Dahlof B, Sever PS, Poulter NR, et al. Prevention of cardiovascular events with an antihypertensive regimen of amlodipine adding perindopril as required versus atenolol adding bendroflumethiazide as required in the Anglo-Scandinavian Cardiac Outcomes Trial-Blood Pressure Lowering Arm (ASCOT-BPLA): a multicenter randomized controlled study. ASCOT investigators. *Lancet*. 2005;366:895–906.
12 Neutel JM, Franklin SS, Oparil S, et al. Efficacy and safety of irbesartan/HCTZ combination therapy as initial treatment for rapid control of severe hypertension. *J Clin Hypertens (Greenwich)*. 2006;8:850–857.
13 Neutel JM, Franklin SS, Lapuerta P, et al. A comparison of the efficacy and safety of irbesartan/HCTZ combination therapy with irbesartan and HCTZ monotherapy in the treatment of moderate hypertension. *J Hum Hypertens*. 2008;22:266–274.
14 Jamerson K, Bakris GL, Dahlöf B, et al. Exceptional early blood pressure control rates: the ACCOMPLISH trial. *Blood Press*. 2007;16:80–86.
15 Beckett NS, Peters R, Fletcher AE, et al. Treatment of hypertension in patients 80 years of age or older. *N Engl J Med*. 2008;358:1887–1898.
16 Mancia G, Laurent S, Agabiti-Rosei E, et al. Reappraisal of European guidelines on hypertension management: a European Society of Hypertension Task Force document. *J Hypertens*. 2009;27:2121–2158.
17 Law MR, Wald NJ, Morris JK, et al. Value of low dose combination treatment with blood pressure lowering drugs: analysis of 354 randomised trials. *BMJ*. 2003; 326:1427–1431.
18 Weir MR, Levy D, Crikelair N, et al. Time to achieve blood-pressure goal: influence of dose of valsartan monotherapy and valsartan and hydrochlorothiazide combination therapy. *Am J Hypertens*. 2007;20:807–815.
19 Neutel JM, Saunders E, Bakris GL, et al. The efficacy and safety of low- and high-dose fixed combinations of irbesartan/hydrochlorothiazide in patients with uncontrolled systolic blood pressure on monotherapy: the INCLUSIVE trial. *J Clin Hypertens (Greenwich)*. 2005; 7:578–586.
20 Cappuccio FP, Kerry SM, Forbes L, et al. Blood pressure control by home monitoring: meta-analysis of randomized trials. *BMJ*. 2004;329:493–499.
21 Staessen JF, Hond ED, Celis H, et al. Antihypertensive treatment based on blood pressure measurement at home or in the physician's office. A randomized control trial. *JAMA*. 2004;291:955–964.
22 Mallion JM, Genes N, Vaur L, et al. Detection of masked hypertension by home blood pressure measurement: is the number of measurements an important issue? *Blood Press Monit*. 2004;9:301–305.
23 Obara T, Ohkubo T, Kikuya M, et al. Prevalence of masked uncontrolled and treated white-coat hypertension

defined according to the average of morning and evening home blood pressure value: from the Japan Home versus Office Measurement Evaluation Study. *Blood Press Monit*. 2005;10:311–316.

24 Kabutoya T, Ishikawa J, Hoshide S, et al. A home blood pressure monitor equipped with a graphic function facilitates faster blood pressure control than the conventional home blood pressure monitor. *J Clin Hypertens (Greenwich)*. 2009;11:422–425.

25 Casalino L. The unintended consequences of measuring quality on the quality of medical care. *N Engl J Med*. 1999;341:1147–1150.

26 Hayward RA. All or nothing treatment targets make bad performance measures. *Am J Managed Care*. 2007; 13:126–128.

JRRD Volume 47, Number 8, 2010
Pages 719–724
Journal of Rehabilitation Research & Development

Can administrative data identify active diagnoses for long-term care resident assessment?

Dan R. Berlowitz, MD, MPH;[1*] Elaine C. Hickey, RN, MS;[1] Debra Saliba, MD, MPH[2]
[1]Center for Health Quality, Outcomes, and Economic Research, Edith Nourse Rogers Memorial Veterans Hospital, Bedford, MA; and Boston University School of Public Health, Boston, MA; [2]Greater Los Angeles Department of Veterans Affairs Geriatric Research, Education, and Clinical Center and Health Sciences Research and Development Center of Excellence, Los Angeles, CA; and University of California Los Angeles/Los Angeles Jewish Homes Borun Center, Los Angeles, CA

Abstract—Many veterans receive rehabilitation services in Department of Veterans Affairs (VA) nursing homes. Efficient methods for the identification of active diagnoses could facilitate care planning and outcomes assessment. We set out to determine whether diagnostic data from VA databases can be used to identify active diagnoses for Minimum Data Set (MDS) assessments. We evaluated diagnoses being considered for inclusion in MDS version 3.0 and present in at least 15% of a sample of VA nursing home residents. A research nurse following a standardized protocol identified active diagnoses from the medical records of 120 residents. A clinical nurse also identified active diagnoses in 58 of these patients. Inpatient and outpatient diagnoses from the VA National Patient Care Database were identified for the past year. We calculated kappa, sensitivity, and specificity values, considering the nurses' assessments the gold standard. We found that kappa values comparing research nurses and databases were generally poor, with only 8 of the 19 diagnoses having a value >0.60. Levels of agreement between the clinical nurse and administrative data were generally similar. We conclude that VA administrative data cannot be used to accurately identify active diagnoses for nursing home residents. How best to efficiently collect these important data remains uncertain.

Key words: active diagnosis, care planning, Community Living Centers, comorbidity, Minimum Data Set, nursing homes, outcomes data, rehabilitation, risk adjustment, veterans.

INTRODUCTION

Rehabilitation for patients with disabilities is increasingly being provided in skilled nursing facilities [1], now known in the Department of Veterans Affairs (VA) as Community Living Centers (CLCs). Critical to assessing and improving the quality of this rehabilitation care is a comprehensive understanding of resident outcomes [2]. Outcomes data may be used to profile CLCs on the quality of their care and to identify benchmarks for best practices within the entire VA. In the examination of outcomes, risk adjustment helps ensure that any observed variations reflect differences in care rather than differences in patient mix. Risk adjustment for rehabilitation outcomes should incorporate many different patient-mix factors, including sociodemographics, functional status,

Abbreviations: CLC = Community Living Center, ICD-9-CM = International Classification of Diseases-9th Revision-Clinical Modification, MDS = Minimum Data Set, TIA = transient ischemic attack, VA = Department of Veterans Affairs.

[*]**Address all correspondence to Dan R. Berlowitz, MD, MPH; Bedford VA Hospital–CHQOER, 200 Springs Road, Bedford, MA 01730; 781-687-2962; fax: 781-687-2227. Email: Dan.Berlowitz@va.gov**

DOI:10.1682/JRRD.2009.08.0123

cognitive ability, and sensory function [3–4]. A number of studies have also shown that comorbidities are an important patient risk factor to consider when adjusting on rehabilitation outcomes [5–7]. Capturing information on comorbidities will then be essential for the development of an outcomes tracking system for VA rehabilitation patients residing in CLCs.

Information on comorbidities is available on all nursing home residents, including those in VA CLCs, through the Minimum Data Set (MDS). This comprehensive resident assessment system was developed in response to the 1986 Institute of Medicine report on improving care in nursing homes [8] and includes information necessary for care planning. Specific sections address topics such as physical function, cognition, behavior, health conditions, and disease diagnoses. However, concerns have long been raised about the use of MDS data for purposes such as quality assessment and research [9–10]. In part, these concerns have been fueled by questions about the reliability of resident assessments, and studies have shown that the correlation among specially trained nurse assessors on various items may be low [11]. The Disease Diagnoses section of the MDS, which contains information on important comorbidities, has been viewed as especially difficult, in part because of the requirement that only active diagnoses be recorded. This requirement reflects the importance of the MDS in care planning, where knowledge of active diagnoses, as opposed to all diagnoses, is critical. Active diagnoses are defined as those that have a relationship to the resident's current functional status, cognitive status, mood or behavioral status, treatments, monitoring plan, or prognosis. The recently completed Data Assessment and Verification project, performed for the Centers for Medicare and Medicaid Services, identified Disease Diagnoses as one of the most common sections for discrepancies, mostly because of diagnoses that were no longer active being recorded in the MDS.

VA has a wealth of diagnostic data in its National Patient Care Database. Because these data are generated from recent hospital, outpatient, or long-term care encounters between patients and clinicians, they may be an alternate source of information on active diagnoses for use on the MDS. Therefore, as part of a validation of the proposed MDS version 3.0, we examined the correlation between VA administrative data and diagnostic data recorded in the MDS. Specifically, for the MDS data, we used MDS assessments performed by both specially trained research nurses and clinical nurses as part of routine care. These results could help inform the accuracy of VA administrative data and whether it may replace assessments currently performed by clinical nurses.

METHODS

Study Setting and Sample

This study was a part of the larger VA MDS 3.0 pilot testing and validation study funded by the Health Services Research and Development Service. Among the many goals of this study was to improve the accuracy of the diagnostic data collected during MDS assessments. Study participants were from four VA CLCs located in the northeast. At each CLC, residents were selected based on their being scheduled for their routine MDS 2.0 assessment, which is typically done on admission, quarterly, and with significant changes in health status. As an additional exclusion criterion, residents could not be comatose.

Minimum Data Set Assessments

Within 48 hours of the required MDS 2.0 assessment, either of two research nurses completed an additional pilot MDS 3.0 assessment. We used this pilot version of MDS 3.0 to collect information on active diagnoses. The Disease Diagnoses section of the pilot MDS 3.0 is similar to that of the currently used MDS 2.0 in terms of the specific diseases captured. However, a major change is the development of more detailed protocols to describe when a disease is active, where in the medical record this information should be sought, and the time frame to be considered for activity. Thus, it stresses first determining whether the condition is present and then whether it is active. As an example, for heart failure, active disease requires a physician-documented diagnosis of heart failure plus one or more of the following: a physician note indicating active disease; a positive test, such as a chest X-ray, within the past 30 days indicating heart failure; signs or symptoms, such as dyspnea, attributed to heart failure; current medication treatment; or hospitalization for heart failure within the past 30 days. Specific International Classification of Diseases-9th Revision-Clinical Modification (ICD-9-CM) codes were assigned to each MDS 3.0 diagnosis to facilitate comparisons with administrative data.

The trained research nurses conducted a detailed review of medical records to identify active diagnoses. These two research nurses had received extensive training in the use of MDS 3.0 and, in the case of the Disease Diagnoses section, had helped in the development of the criteria used to determine disease activity. Thus, the research nurses may be considered as the "gold standard" assessment. A total of 120 patients were evaluated by the research nurses.

Fifty-eight of these patients also had a pilot MDS 3.0 assessment completed by the clinical team. Typically, these assessments were performed by the MDS coordinator on the unit and were based on the assessor's knowledge of the resident, discussions at team meeting, and review of the medical records. These nurse-assessors had received more limited training in the use of the instrument and could be considered to represent how the assessment would typically be completed in actual clinical practice.

Department of Veterans Affairs Administrative Data

We used the VA National Patient Care Database to collect all ICD-9-CM codes from the year before the MDS 2.0 assessment for the 120 patients. We used ICD-9-CM codes from hospital, outpatient, and long-term care settings. However, we excluded codes from nonclinician visits, such as laboratory or radiology. No diagnostic data were collected from non-VA sources such as Medicare.

Analyses

Separate analyses were performed for the research and clinical nurses. We examined those MDS 3.0 diagnoses present in at least 15 percent of the patients when assessed by any source, whether research nurse, clinical nurse, or administrative data. Two-by-two tables were constructed for the presence or absence of each diagnosis in the nurse assessment and in the VA administrative data. Overall level of agreement between the two data sources was calculated for each diagnosis with use of the kappa statistic. Sensitivity and specificity were then calculated, with the nursing assessment as the gold standard. Thus, sensitivity described what proportion of patients identified by the nurse as having the disease was also identified as having the disease in the administrative data and specificity described what proportion of patients identified by the nurse as not having the disease was also identified as not having the disease in the administrative data.

RESULTS

Nineteen diagnoses were evaluated. For most diagnoses, limited agreement existed between the research nurses and the administrative data (**Table**). In only eight diagnoses did the kappa value equal or exceed 0.60: uncomplicated diabetes mellitus, stroke/transient ischemic attack (TIA), coronary artery disease, chronic heart failure, thyroid disorder, hemiplegia/paraplegia/quadriplegia, asthma/chronic obstructive pulmonary disease, and schizophrenia. For other diagnoses, the level of agreement was generally poor, with a kappa level as low as 0.18 for depression. Results were very similar when clinical nurses were compared with administrative data (**Table**), again with only eight of the diagnoses having kappa values exceeding 0.60. Research and clinical nurses were also similar in terms of which diagnoses they had high and low levels of agreement on with administrative data. The eight diagnoses with the highest kappa values (>0.60) for the research nurses included six diagnoses with the highest kappa values for the clinical nurses. The six diagnoses with low kappa values for the research nurses (<0.40) included the four with the lowest kappa values for the clinical nurses. No clear pattern was evident as to which diagnoses had high or low levels of agreement. Mental health disorders included the diagnoses with the highest kappa value, schizophrenia, and the lowest kappa value, depression.

Sensitivity of administrative data compared with the research nurses varied considerably, ranging from 30 percent for depression to 100 percent for both stroke/TIA and hemiplegia/paraplegia/quadriplegia. Low sensitivity indicates that diagnoses identified by the research nurse as present may not be listed in the administrative data. Specificity, in most cases, was better than sensitivity and varied less, with a range from 77 to 100 percent. High specificity indicates that the administrative data rarely listed diagnoses as present within the past year when the nurses indicated it was absent. Results from the clinical nurses were generally similar to the research nurses.

DISCUSSION

Accurate information on diagnoses is essential in care planning and tracking of outcomes of nursing home residents receiving rehabilitation. Numerous studies have confirmed the validity of the diagnostic data contained in

Table.
Comparisons between administrative data and research and clinical nurses' identification of nursing home residents' active diagnoses. In calculating sensitivity and specificity, we considered nurses as "gold standard."

Diagnosis in Administrative Data	Research Nurse			Clinical Nurse		
	Kappa	Sensitivity (%)	Specificity (%)	Kappa	Sensitivity (%)	Specificity (%)
Arrhythmias (nonatrial fibrillation)	0.31	57	90	0.38	100	86
Coronary Artery Disease	0.65	68	94	0.72	80	91
Chronic Heart Failure	0.60	68	93	0.56	78	88
Hypertension	0.47	74	77	0.46	89	55
GERD/Peptic Ulcer	0.48	54	91	0.68	65	100
Benign-Prostatic Hypertrophy	0.52	57	93	0.59	64	94
Anemia	0.36	48	87	0.38	45	93
Uncomplicated Diabetes Mellitus	0.69	78	91	0.65	100	80
Arthritis	0.36	53	87	0.53	60	91
Stroke/TIA	0.77	100	93	0.90	92	98
Hemiplegia/Paraplegia/Quadriplegia	0.91	100	97	0.71	89	92
Dementia: Alzheimer Disease	0.47	57	96	0.62	100	91
Dementia: Non-Alzheimer Disease	0.37	34	97	0.31	36	92
Asthma/COPD	0.63	68	94	0.63	59	98
Cancer	0.35	80	83	0.59	80	84
Thyroid Disorder	0.67	55	100	0.46	50	96
Anxiety Disorder	0.50	55	95	0.48	50	94
Depression	0.18	30	87	0.20	33	85
Schizophrenia	0.94	94	99	0.85	100	96

COPD = chronic obstructive pulmonary disease, GERD = gastroesophageal reflux disease, TIA = transient ischemic attack.

VA administrative records [12–14]. However, for nursing homes, the MDS requires that the diagnosis be not only present but active. This added requirement has not been previously examined in the VA. Given the difficulties clinical staff have in identifying active diagnoses, we hypothesized that VA administrative data might serve as a useful substitute in the completion of MDS for CLC residents.

Our results did not support this hypothesis. We found that the level of agreement, as reflected by kappa values, was generally low when we compared administrative data and research nurses. Kappa values were greater than or equal to 0.60 for only 8 of the 19 conditions. Tremendous variability was also found in sensitivity and specificity of administrative data, although specificities were generally higher. This suggests that when diagnoses are listed in VA administrative data within the past year, they do reflect conditions that are active. Using a time frame longer than 1 year would be expected to increase the sensitivity and reduce this specificity.

Results were not substantially different when we compared clinical nurses and administrative data. Levels of agreement were often poor, and for only eight conditions was the kappa greater than 0.60. However, perhaps reflecting the fact that information in administrative databases is generally derived from clinicians, specificity remained high.

Relatively few studies have examined whether administrative databases could replace clinicians' assessments on the MDS. In one study of Ontario nursing homes, the MDS often did not include many important diagnoses that had been present in the discharge diagnosis database from the preceding hospitalization [15]. Reasons for these discrepancies were unclear but thought to possibly reflect the incomplete transfer of diagnostic information upon resident transfer between settings. Other studies have either compared different research nurses [11] or compared clinicians' MDS assessments with trained assessors and standard protocols [16–17].

VA administrative data are generally felt to be more comprehensive than databases from other healthcare settings. Thus, it is difficult to imagine that other databases would be better able to identify active diagnoses. Results from this study, then, would be applicable to MDS assessments outside the VA. However, our study did not use the additional diagnostic data available in Medicare

files. Studies have shown that incorporation of Medicare data improves the capture of comorbidity burden in veterans who are dual users. The addition of this diagnostic data contained in Medicare files would be expected to increase sensitivity but reduce specificity; the effect on kappa values would be uncertain. Further studies would be required to determine whether additional diagnoses from Medicare would assist in the accurate identification of active diagnoses within the VA.

Several additional limitations of this study should be noted. We only examined four nursing homes located in the northeast. Results could differ in other locations. Our sample size was also relatively small, so the number of patients per diagnosis was low. Furthermore, several diagnoses from MDS 3.0 were excluded because they were present in less than 15 percent of the sample. We do not know whether administrative data would be better at coding these rare conditions.

While study results highlight that administrative data should not be used for the identification of active diagnoses on the MDS, our results do not suggest how the identification of active diagnoses may be improved. Our assumption is that the research nurses most accurately identified active diagnoses because of their reliance on strict protocols of medical record reviews. Additional training of clinical nurses on completion of the MDS would then be required to ensure the most accurate information on active diagnoses in nursing home residents. However, research nurses could miss important diagnoses because of poor documentation. The Ontario study suggests that improved transfer of data from hospital stays could help improve the identification of MDS diagnoses [15]. Given VA's electronic medical records, we believe a lack of data on transfer is a less likely explanation for our results. Additional studies are clearly indicated.

CONCLUSIONS

The MDS is a valuable tool for VA clinicians, managers, and researchers working with rehabilitation patients in CLC settings. An important aspect of the MDS is the Disease Diagnoses section that provides information essential in care planning and outcomes measurement. Despite the importance of these data, studies have shown that clinical staff poorly identify active diagnoses when completing the MDS [16]. Our results suggest that administrative data cannot substitute for the assessments currently performed by VA clinicians.

ACKNOWLEDGMENTS

Author Contributions:
Study concept and design: D. R. Berlowitz, E. C. Hickey, D. Saliba.
Acquisition of data: E. C. Hickey.
Analysis and interpretation of data: D. R. Berlowitz, E. C. Hickey, D. Saliba.
Drafting of manuscript: D. R. Berlowitz.
Critical revisions of manuscript: E. C. Hickey, D. Saliba.
Obtained funding: D. R. Berlowitz.
Financial Disclosures: The authors have declared that no competing interests exist.
Funding/Support: This material was based on work supported by the VA, Veterans Health Administration, Health Services Research and Development Service (grant SDR 03-217).
Institutional Review: The study protocol was approved through the VA Human Subjects Protection and Privacy Board process.

REFERENCES

1. Retchin SM, Brown RS, Yeh SC, Chu D, Moreno L. Outcomes of stroke patients in Medicare fee for service and managed care. JAMA. 1997;278(2):119–24. [PMID: 9214526] DOI:10.1001/jama.278.2.119
2. Rao P, Boradia P, Ennis J. Shift happens: Using outcomes to survive and thrive under PPS. Top Stroke Rehabil. 2005; 12(2):1–3. [PMID: 15940579] DOI:10.1310/UG5D-4RBT-VET6-AJ29
3. Iezzoni LI. Risk adjusting rehabilitation outcomes: An overview of methodologic issues. Am J Phys Med Rehabil. 2004;83(4):316–26. [PMID: 15024335] DOI:10.1097/01.PHM.0000118041.17739.BB
4. Kane RL. Improving outcomes in rehabilitation. A call to arms (and legs). Med Care. 1997;35(6 Suppl):JS21–27. [PMID: 9191711] DOI:10.1097/00005650-199706001-00004
5. Hoenig H, Sloane R, Horner RD, Zolkewitz M, Reker D. Differences in rehabilitation services and outcomes among stroke patients cared for in veterans hospitals. Health Serv Res. 2001;35(6):1293–1318. [PMID: 11221820]
6. Ottenbacher KJ, Smith PM, Illig SB, Linn RT, Ostir GV, Granger CV. Trends in length of stay, living setting, functional outcome, and mortality following medical rehabilitation. JAMA. 2004;292(14):1687–95. [PMID: 15479933] DOI:10.1001/jama.292.14.1687
7. Berlowitz DR, Hoenig H, Cowper DC, Duncan PW, Vogel WB. Impact of comorbidities on stroke rehabilitation

outcomes: Does the method matter? Arch Phys Med Rehabil. 2008;89(10):1903–6. [PMID: 18929019] DOI:10.1016/j.apmr.2008.03.024

8. Institute of Medicine Committee on Nursing Home Regulation. Improving the quality of care in nursing homes. Washington (DC): National Academy Press; 1986.
9. Teresi JA, Holmes D. Should MDS data be used for research? Gerontologist. 1992;32(2):148–49. [PMID: 1577305]
10. Brooks S. What's wrong with the MDS (Minimum Data Set)? Contemp Longterm Care. 1996;19(11):41,43,45–47. [PMID: 10162289]
11. Hawes C, Morris JN, Phillips CD, Mor V, Fries BE, Nonemaker S. Reliability estimates for the Minimum Data Set for nursing home resident assessment and care screening (MDS). Gerontologist. 1995;35(2):172–78. [PMID: 7750773]
12. Kashner TM. Agreement between administrative files and written medical records: A case of the Department of Veterans Affairs. Med Care. 1998;36(9):1324–36. [PMID: 9749656] DOI:10.1097/00005650-199809000-00005
13. Szeto HC, Coleman RK, Gholami P, Hoffman BB, Goldstein MK. Accuracy of computerized outpatient diagnoses in a Veterans Affairs general medical clinic. Am J Manag Care. 2002;8(1):37–43. [PMID: 11814171]
14. Borzecki AM, Wong AT, Hickey EC, Ash AS, Berlowitz DR. Identifying hypertension-related comorbidities from administrative data: What's the optimal approach? Am J Med Qual. 2004;19(5):201–6. [PMID: 15532912] DOI:10.1177/106286060401900504
15. Wodchis WP, Naglie G, Teare GF. Validating diagnostic information on the Minimum Data Set in Ontario Hospital-based long-term care. Med Care. 2008;46(8):882–87. [PMID: 18665069]
16. Centers for Medicare & Medicaid Services [Internet]. MDS 2.0 for nursing homes. Baltimore (MD): Centers for Medicare & Medicaid Services. [updated 2010 Mar 29; cited 2009 Aug 5]. Available from: http://www.cms.gov/nursinghomequalityinits/20_nhqimds20.asp
17. Stevenson KB, Moore JW, Sleeper B. Validity of the Minimum Data Set in identifying urinary tract infections in residents of long-term care facilities. J Am Geriatr Soc. 2004; 52(5):707–11. [PMID: 15086649] DOI:10.1111/j.1532-5415.2004.52206.x

Submitted for publication August 13, 2009. Accepted in revised form December 29, 2009.

This article and any supplementary material should be cited as follows:

Berlowitz DR, Hickey EC, Saliba D. Can administrative data identify active diagnoses for long-term care resident assessment? J Rehabil Res Dev. 2010;47(8):719–24.

DOI:10.1682/JRRD.2009.08.0123

Risk Adjustment in Rehabilitation Quality Improvement

Dan R. Berlowitz, MD, MPH,[1,2] and Margaret Stineman, MD[3]

[1]Center for Health Quality, Outcomes and Economic Research, Bedford VA Hospital, Bedford, Massachusetts; [2]Boston University Schools of Public Health and Medicine, Boston, Massachusetts; [3]Department of Physical Medicine and Rehabilitation and the Center for Epidemiology and Biostatistics, University of Pennsylvania, Philadelphia, Pennsylvania

Risk adjustment is an important tool for examining quality and outcomes of care. Considerable work has been accomplished in developing and applying risk adjustment to a variety of health care settings. The knowledge gained from these experiences is increasingly available to help guide clinicians, managers, and policy makers in the area of stroke rehabilitation. Key issues include understanding the purpose of risk adjustment, the unique population being examined, and the specific outcomes of interest. Risk adjustment should also be based on models that have good statistical properties and also make sense to clinicians working in the field. Risk-adjustment systems such as the FIM™- function-related groups (FIM-FRGs) are now being used to monitor the quality of stroke rehabilitation. A specific application of this risk-adjustment system is demonstrated. **Key words:** *quality improvement, quality of health care, risk adjustment, stroke rehabilitation*

Clinicians, managers, and policy makers working in the area of stroke rehabilitation increasingly must address problems that require a comprehensive understanding of patient outcomes. Clinicians are interested in knowing about the quality of their care and whether observed outcomes are better than expected. Managers need to develop quality improvement initiatives and ensure that there are sufficient resources to care for their patients. Policy makers seek new methods to reduce costs and reward clinicians who provide optimal care. To achieve all these things, the characteristics, or case mix, of the population of interest needs to be defined. Without this information, it is often difficult to know whether poor outcomes in a group of patients are due to poor care or to a sicker population.

In the past 30 years, there have been tremendous advances in the knowledge of risk adjustment, the process by which the case mix of patients is described and the information used for predicting or understanding diverse outcomes. The central element of risk adjustment is determining expected outcomes based on intrinsic patient characteristics. To accomplish this, a risk-adjustment model is used that was developed for a specified large group or population, such as all patients in the country with a stroke. The model is then used to determine expected outcomes in a smaller population of interest, such as stroke patients at a particular hospital. Actual outcomes at the hospital may then be compared to these expected outcomes. The outcomes themselves may be clinical outcomes such as mortality or functional recovery. Alternatively, the outcomes may be measures of resource utilization such as expected costs or need for hospitalization. The process by which risk adjustment is performed has been well-defined, and numerous systems have been developed for use in specific circumstances. This large body of knowledge is now available to guide those working in the area of stroke rehabilitation.

What Is Risk Adjustment?

At its simplest, risk adjustment is a process that allows us to compare "apples and oranges." To make fair comparisons of different providers, treatments, or populations on an outcome of interest, we must take into account that some groups will have sicker patients than others. As outcomes studies rarely rely on randomizing subjects to different interventions, risk adjustment is a way to reduce the affects of these confounding factors that may influence the outcome of interest. It is the means of accounting for those patient factors that are

Top Stroke Rehabil 2010;17(4):252–261

doi: 10.1310/tsr1704-252

likely to vary among providers and also affect the outcome of interest. Many different terms are used to describe these patient factors including case mix, patient mix, disease severity, comorbidity, health status, and complexity. Although there are subtle differences in the definitions of many of these terms, they all are similar in conveying a level of risk for a certain outcome.

In considering risk-adjustment approaches for stroke rehabilitation, or any other purpose, Iezzoni[1] has emphasized a series of questions that should be addressed. First is "risk for what?" Approaches to risk adjustment will vary depending on the outcome of interest. Typically, risk adjustment has focused on outcomes such as mortality, cost, or hospitalizations. However, in stroke rehabilitation, functional outcomes such as physical and cognitive performance may be especially important. Thus, the outcome of interest should be explicitly specified.

A second question is "over what time frame?" Outcomes of interest and their predictors are likely to vary dramatically depending on the length of time people are being followed. A simple change, such as looking at 30-day mortality rather than hospital mortality, can alter what is found.[2]

The third question is "for what population?" Different groups will have very different risks for certain outcomes. In looking at mortality, children are obviously very different than adults, whereas the risks of developing many diseases are very different for women and men. In the area of stroke rehabilitation, it may be extremely important to differentiate between ischemic and hemorrhagic strokes, and outcomes, as well as their predictors, will vary in these 2 subpopulations.

The final question is "for what purpose?" When risk adjustment is used to evaluate individual clinicians, especially if important decisions such as credentialing or performance payments are at stake, it is critical that the risk adjustment be credible both from a clinical and statistical perspective. If important dimensions of risk are not captured, clinicians can excuse their poor outcomes by saying that their patients are sicker. In contrast, for outcomes such as capitated payments for health plans, risk-adjustment approaches with only limited statistical performance will often be sufficient.

Risk adjustment should strive to account for all important dimensions of risk. These dimensions of risk could include a wide range of patient-specific characteristics that are associated with an increased (or decreased) risk for the outcome of interest.[1,3] Demographics such as age, sex, and race/ethnicity may be included. Clinical factors might include the principal diagnosis, comorbidities, disease-specific severity, and physiologic derangements. In examining rehabilitation outcomes, baseline functional status is likely to be particularly important. Socioeconomic status, health-related behaviors, and specific attitudes may also be important dimensions of risk. Which of these factors are included in risk adjustment will often depend on the outcome of interest, the range of factors that are clinically believed to be important predictors of that outcome, and the availability of the data of interest. Factors such as age, gender, and comorbidities are often easily available in existing administrative databases, but other dimensions of risk may be more difficult to capture.

A clear clinical conceptualization of the association between individual risk factors and the outcome of interest should always guide risk adjustment. For example, outcomes often vary between different racial/ethnicity groups. Before including race/ethnicity as a risk factor, though, it must be clear as to whether the difference in outcomes is potentially due to inherent genetic factors in different racial/ethnic groups or to external factors such as bias in the health care system. In the later case, inclusion of race/ethnicity during risk adjustment would essentially excuse those providing poor care to certain racial/ethnic groups.[1] Risk adjustment prevents providers from being penalized for worse outcomes among patients with the selected dimensions of risk.

How Is Risk Adjustment Performed?

The simplest approach to risk adjustment would be a stratification in which patients are divided into 2 groups based on a single patient characteristic. The characteristic may be age or type of stroke and would be known to have a large impact on the outcome of interest. Within each stratum, patients would be compared on their outcomes. Although simple, the problem with this approach lies in the multidimensional nature of

risk. There are many different factors that need to be considered when predicting risk for a particular outcome. With stratification, only a single factor is being considered and all others are ignored.

Consequently, multivariate regression models that consider the independent effects of many different risk factors on the outcome are typically used in risk adjustment. These models are developed in very large groups of patients through a process that involves a close collaboration between clinicians and statisticians. Clinicians initially identify potential risk factors based on general medical knowledge and literature reviews. Both bivariate and multivariate statistical techniques are then used to test the associations of the potential risk factors with the outcome. Ideally, the results of this highly iterative process will be a risk-adjustment model that is both clinically credible and has good statistical properties. Based on the model coefficients, an expected rate of the outcome is calculated for each patient.

While the ideal approach to any risk-adjustment project may be to develop a model that is unique to the exact population and outcome being examined, the reality is that this is a very difficult and time-consuming process.[4,5] Consequently, for most purposes, it is not feasible to develop a new model. Instead, it is better to rely on an existing risk-adjustment approach that has been developed for a similar outcome. These "off-the-shelf" systems are often easy to apply and are a much better alternative than the option of not doing risk adjustment. It must be recognized that because they were developed in a different population, these systems are not likely to perform as well in the new setting of interest.[6] Any off-the-shelf system that is being considered for use should be carefully reviewed for purpose of development, variety of previous applications, and its ability to capture important dimensions of risk. Simple fixes to existing systems, such as adding a few variables or recalibrating the model, can often dramatically improve their performance in a new setting.

Examples of Existing Risk-Adjustment Systems

A wide variety of risk-adjustment systems have been developed for many different purposes. Perhaps the most widely used system is diagnosis-related groups (DRGs) that aims to predict hospital costs based on relatively few predictors including the primary diagnosis and age.[7] In predicting cost for nursing homes, resource utilization groups (RUGs) consider a broader group of predictors including functional status.[8] For clinical outcomes, the Acute Physiology and Chronic Health Evaluation (APACHE) has been widely used to predict mortality in intensive care unit patients based on both specific conditions present as well as the extent to which physiologic markers such as blood pressure, pulse, and electrolytes are abnormal.[9]

Some of the most extensive work has been the development of risk-adjustment systems that take advantage of information on comorbidity burden. The Charlson Index has been particularly widely used; it considers 16 different serious comorbidities and assigns weights based on the presence and, in a few situations, the severity of the condition.[10] Although originally developed for predicting 1-year mortality among women with breast cancer, it has been used for a variety of purposes including as a risk adjustor when length of stay and discharge destination for stroke rehabilitation patients have been examined.[11] Increasingly, health services research studies have relied on the Elixhauser measure that provides a method for grouping clinically similar diagnoses into 30 distinct categories, each of which can be examined for associations with the outcome of interest.[12] Several systems are also being used that were originally developed to predict resource utilization and costs among outpatients. These include adjusted clinical groups (ACGs), which group conditions similar on characteristics that include persistence of the condition, likelihood of requiring hospitalization, and need for specialty referral, and diagnosis cost groups (DCGs), which group conditions into clinical hierarchical clusters.[13,14] One study found that both ACGs and DCGs were better at predicting change in functional status, rehospitalization, and death among stroke rehabilitation patients than models that relied only on age and gender or the Charlson Index.[15]

In the area of stroke rehabilitation, the FIM™* function-related groups (FIM-FRGs) is the most widely used risk adjustment system.[16] Analogous

*FIM™ is a trademark of Uniform Data System for Medical Rehabilitation, a division of UB Foundation Activities, Inc.

to the DRGs and RUGs, the FIM-FRGs were initially designed specifically to characterize patients undergoing inpatient rehabilitation for reimbursements. FRGs include specific measures for 21 primary diagnostic reasons for rehabilitation, 1 of which is stroke. Within each impairment category, the system divides patients into groups according to the severity of their physical and cognitive disabilities as well as age at the time of admission to rehabilitation. The severity of initial physical and cognitive disability is expressed by the FIM™ motor and cognitive subscales.[17] The motor FIM™ includes patients' clinician-rated levels of performance on 13 physical activities, whereas the cognitive FIM™ accounts for 5 communication and social cognition activities. Higher motor and cognitive FIM™ scores are associated with less severe physical and cognitive disabilities, respectively.

Lessons Learned From Nonrehabilitation Settings

The extensive work done with risk adjustment in nonrehabilitation settings has highlighted many important lessons. Risk adjustment must be clinically credible. Clinicians must be able to look at what patient characteristics are being adjusted for and agree that they make sense. To accomplish this, clinically detailed information is often required. Such clinical detail needs to be tailored to the setting of interest. In long-term care, functional status may be key; whereas in the intensive care unit, physiologic derangements are often more important than specific diagnoses. The use of such clinically detailed information in risk adjustment will often be difficult as it is rarely available in databases. Studies have demonstrated, though, that risk-adjustment models incorporating only a limited number of clinically detailed variables can perform almost as well as models with more complete information. For example, the addition of laboratory results to a risk-adjustment model for in-hospital mortality led to substantial improvements in model performance.[18] Subsequent addition of vital signs and other key clinical findings resulted in only modest further improvements. Increasingly, the availability of electronic medical records will likely solve the need to easily obtain detailed clinic information on large numbers of patients.

Not only are unadjusted outcomes data not clinically credible, they are also unlikely to provide accurate assessments of performance. Studies from many settings have shown that providers' performance rates may vary dramatically depending on whether unadjusted or risk-adjusted rates are used.[5,19–21] Furthermore, which providers are identified as outliers will also change considerably. Thus, using unadjusted data in quality improvement initiatives may convey false assessments.

Use of off-the-shelf risk-adjustment systems also can be problematic. Even though they all purport to be measuring severity, different risk-adjustment systems will often come up with very different conclusions.[21,22] For individual patients, the predicted probability of the outcome of interest may vary considerably depending on the system used. In one study of in-hospital mortality among 9,407 stroke patients at 94 hospitals, 5 risk-adjustment systems were compared.[22] Expected probability of death was calculated in every patient using each system. When one system was compared to another one, they could differ considerably in over 60% of patients on the expected probability of death. When applied to individual providers, these systems would often identify different hospitals as outliers. In fact, in identifying outliers, any individual system is often more likely to agree with unadjusted performance than with another risk-adjustment system. Considerable care is required in deciding which off-the-shelf system should be used.

Lessons Learned From the Rehabilitation Setting

The extensive work done with risk adjustment in the rehabilitation setting leads to important lessons that parallel those from nonrehabilitation settings, but it also highlights the distinct qualities of the rehabilitation process. Risk adjustment must be clinically credible, reflecting differences in the unique philosophies and objectives of rehabilitation when compared to the acute and long-term care phases of care as well as clear differences in the population served.[1] Rehabilitation

professionals need to be assured that the patient characteristics being adjusted for make sense to their therapeutic objectives and will reflect expected patient differences in essential outcomes such as functional outcomes and discharge to the least restrictive environment. As in other settings, to accomplish this, clinically detailed and particular types of information will be required. Patients' initial functional status is a key measure of severity and need for resources, as in long-term care, but so is diagnostic information about the specific condition for which rehabilitation services are being provided and comorbidities, variables more seminal to case-mix measurement in acute care. Certainly, the amount of therapy needed and the outcomes expected will differ for stroke patients who are older, with more physical and cognitive disabilities, than for those who are younger and less disabled. Clearly, rehabilitation targeted to persons with stroke differs from that targeted to those with spinal cord injury or some other condition, and the need for medical management during rehabilitation will be greater among persons with major comorbidities such as end-stage renal disease than among those who are otherwise healthy. Finally, with outcomes such as home discharge, the qualities of patients' home environments including architecture, and the availability of caregivers are essential determinants. Such ecological information is rarely captured in medical approaches to case-mix measurement. Obtaining the breadth of this clinically detailed information in risk adjustment is difficult. It is rarely available in administrative records, and the analysis of all pertinent dimensions would depend on multivariable regression models that require statistical and analytic expertise, which is not easily available to most clinicians. As in other settings, studies in rehabilitation demonstrate that risk-adjustment models incorporating only a few of the most essential clinical variables can perform almost as well as models with more complete information.

Risk-adjustment systems such as FIM-FRGs,[15] which were initially established to scale reimbursement to case-mix complexity, have been shown to be useful in large and sustained quality improvement projects on stroke rehabilitation in the United States and in other nations. For one US example, the Veteran's Health Administration (VHA) embedded FIM-FRGs in its national treatment algorithm for monitoring stroke rehabilitation across its continuum of postacute rehabilitation services.[23] The algorithm, developed by a national expert panel within the VHA, was intended to promote access to care by advocating for the functional assessment of stroke patients across all severity groups and to encourage the early initiation of treatment and smooth transitions across levels of rehabilitative care. Within that algorithm, FRGs accounted for differences in patient characteristics across alternative levels of rehabilitation and were used to monitor and project resource use and outcomes. Expectations for functional outcome and total treatment costs were found to be similar within FRG groups regardless of whether patients underwent rehabilitation in medical or surgical units, inpatient rehabilitation units, or outpatient settings.

The FRG system also was effective in addressing differences in rehabilitation outcomes and resource use associated with structural variation. In Canada, where people have universal access to publically funded programs for stroke rehabilitation, the Canadian Institute for Health Improvement established FRG-specific benchmarks for stroke rehabilitation and applied an ongoing outcome pattern analytic approach.[24] Pattern analytic approaches work by comparing outcomes and resource use distributions across clinically similar groups of people in ways that make it possible to compare the outcomes of alternative health care systems or approaches.[25] FRGs proved to be a way to define clinically similar groups. For example, over a 6-year period, when examined within FRGs, more than 75% of patients in Canada had longer rehabilitation stays than what was considered a typical stay in the United States. The 75th percentile FIM™ discharge score ratings were higher for all and the 25th percentile FIM™ score ratings were higher for most FRGs in Canada than in the United States. The authors suggested that stroke patients may be discharged in the United States before they achieve optimal functional recovery. Comparisons of VHA to non-VHA resource use and outcomes within the United States showed similar pattern differences with VA average length of stays 30% to 200% longer

by FRG and average functional recoveries better in most FRGs. We now illustrate a simple pencil and paper approach to case-mix adjusted quality improvement applying a FIM-FRG algorithm that provides the means to adjust for patient complexity using 4 of the most essential clinical variables as noted previously.

A Case-Mix Adjustment Strategy Using a Clinical Vignette

For this example, we selected the vignette, Maintaining Functional Outcomes in Times of Change, from the introductory article by Strasser. To show the nuts and bolts of case-mix–adjusted continuous quality improvement (CQI), we selected the patient's motor FIM™ score achieved at rehabilitation discharge as the outcome of interest. The technique applies an outcome pattern analysis approach that looks at expected outcomes as falling within lower, middle, and higher ranges according to high gold standard FRG-adjusted benchmarks. The benchmarks reflect expected values calculated from a risk-adjustment model derived in over 55,000 patient discharges in 1995 from 467 rehabilitation facilities in the United States. They are considered to represent optimal patterns of functional recovery, because they were calibrated to reflect a time before the potentially distorting effects of the current rehabilitation prospective payment system. Stability of the benchmarks is further supported by the demonstration of remarkable consistency with FRG-adjusted values obtained from earlier years (1990 and 1992). Because of changes in the health care system and cost-containment pressures, these benchmarks might not be reasonably achieved today. They are intended to represent a high or optimal standard of outcome achievement, goals to shoot for in efforts to help patients achieve the highest possible functional outcomes after stroke.

The FIM-FRGs classification includes a series of separate modular measures calibrated to establish expected resource use and outcomes. The particular FRG module used here was designed to group patients according to their expected motor FIM™ outcome achievements at discharge from rehabilitation. Compared to the length of stay (LOS) and cost-predicting modules, the discharge motor outcome tool (DMF-FRG) requires more patient groups to adequately distinguish patients by expected functional outcomes.[26] **Figure 1** displays the DMF-FRG functional outcome benchmarks for stroke rehabilitation. The median expected discharge motor FIM™ score for each DMF-FRG is indicated by the middle horizontal line. The gold standard benchmark is a range of scores from the 25th to the 75th percentile. The lower horizontal line represents the 25th percentile lower limit of the motor FIM™ score benchmark and the upper horizontal line forms the 75th percentile upper limit of the benchmark. Half of all patients would be expected to have scores within the benchmark, that is, between the lower and upper lines according to the 1995 gold standard.

Comparison of the FRG definitions in **Table 1** to the pattern of plotted benchmarks in **Figure 1** graphically depicts how a patient's admission motor FIM™, cognitive FIM™, and age interact to adjust risk. Admission motor FIM™ score is the dominant variable in determining functional outcome as represented by the upward diagonal pattern of 25th and 75th percentile benchmarks. The expected discharge motor FIM™ score increases directly with the patient's admission score such that higher admission motor FIM™ scores are associated with the achievement of higher motor FIM™ discharge scores. Comparison of outcome expectations across the FRGs shows the degree to which it would be invalid to compare FIM™ motor discharge scores across populations of stroke patients presenting with different levels of functional severity. The downward trends from the dominant diagonal line, for example, in the benchmarks associated with FRGs 4, 5, and 6, show the negative prognostic effects of advanced age on functional achievements holding initial physical disability constant. The negative effects of advanced age appear most prominent among persons presenting with the most severe physical disabilities (lowest motor FIM™ scores). Among those with milder physical disabilities, cognitive dysfunction has a negative impact on prognosis interacting with both physical and cognitive disability at FRG 14 and 15. Because this article focuses on case-mix

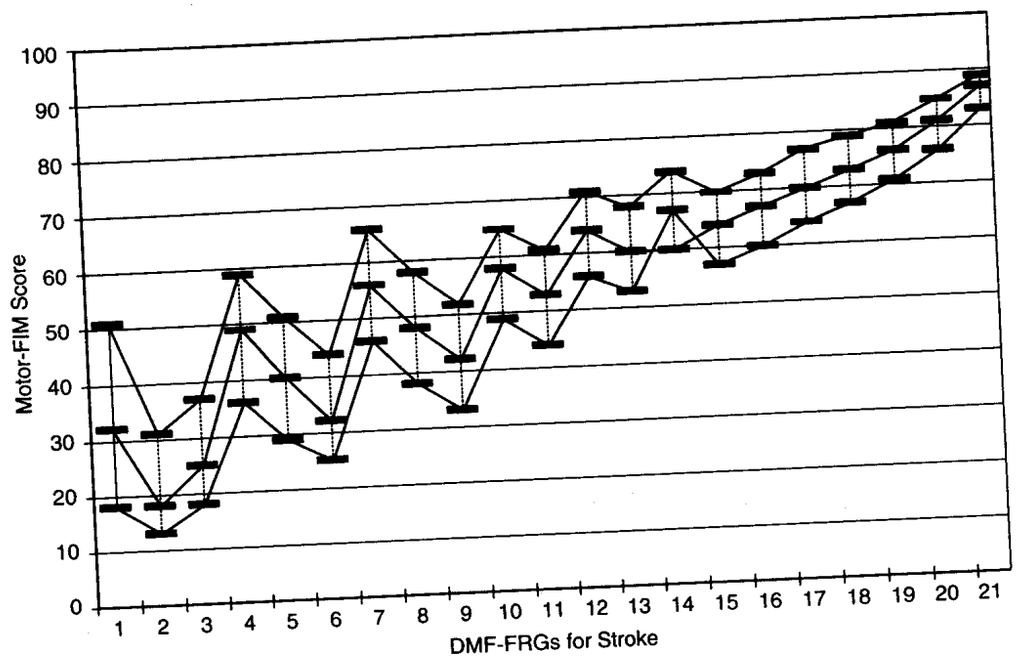


Figure 1. Discharge motor FIM™ benchmark gold standard. The figure describes discharge motor FIM™ function-related groups (DMF-FRG) (x-axis) and functional outcome benchmarks (y-axis) for stroke rehabilitation. These can be used in case-mix–adjusted continuous quality improvement (CQI) once all patients discharged over a specified period are assigned to DMF-FRGs. Each DMF-FRG–specific set of benchmarks defines a range of discharge motor FIM™ scores spanning the 25th through the 75th percentile. The lower horizontal line represents the 25th percentile lower limit of the benchmark and the upper horizontal line forms the 75th percentile upper limit of the benchmark. Percentiles were calculated from the actual outcomes achieved by over 34,000 patients treated across the United States in 1995 whose stroke rehabilitation occurred prior to implementation of Medicare's inpatient rehabilitation facility prospective payment system. To perform case-mix–adjusted CQI determine the proportion of patients each cycle whose functional outcomes are below, within, and above the DMF-FRG–specific benchmark in a particular cycle. Changes in proportions can be tracked over time through subsequent CQI cycles in efforts to address the impact of specific strategies designed to enhance patients' functional gains. Reproduced, with permission, from Stineman MG, Granger CV. Outcome, efficiency, and time-trend pattern analyses for stroke rehabilitation. *Am J Phys Med Rehabil.* 1998;77:193–201. Copyright © 1998 by Lippincott Williams & Wilkins.

adjustment, we use the vignette to illustrate how to case-mix adjust functional outcomes during the evaluate phase of the quality improvement cycle (see vignette, Maintaining Funtional Outcomes in Times of Change, from the introductory article by Strasser). A more detailed example of case-mix adjusted quality improvement applying this same vignette will be given in a separate article (see Stineman et al in this issue). That article is intended to illustrate implementation of a more complicated multidimensional case-mix adjustment strategy within the engage, educate, and execute phases of the full Quality Improvement/Knowledge Translation model as established by Pronovost and coworkers[27] and detailed by Needham and Korupolu in this issue.

This quality improvement initiative is intended to track and test ongoing therapeutic strategies to maintain motor FIM™ outcomes in the overall stroke population as structural changes imposed on this health system in this vignette are shifting stroke patients from higher intensity to lower intensity rehabilitation services. The FIM-FRG case-mix adjustment allows fairer comparisons across the functionally heterogeneous population of stroke patients seen in treatment and provides a way to compare the functional outcomes achieved in the inpatient rehabilitation facility (IRF) and

Table 1. Assigning the patient's stroke-specific DMF-FRG

Discharge FRG	Admission motor FIM™ score range	Admission cognitive FIM™ score range	Age range
1	18–16	-	13–63
2	13	-	>63
3	14–16	-	>63
4	17–22	-	16–58
5	17–22	-	59–73
6	17–22	-	>73
7	23–28	-	16–52
8	23–28		53–76
9	23–28	-	>76
10	29–34	-	16–70
11	29–34	-	>70
12	35–40	-	16–67
13	35–40	-	>67
14	41–48	5–20	16–76
15	41–48	5–20	>76
16	41–43	21–35	-
17	44–48	21–35	-
18	49–58	5–28	-
19	49–58	29–35	-
20	59–68	-	-
21	70–91	-	-

Note: This table can be used for classifying stroke patients into their discharge motor FIM-FRG (DMF-FRG) on the basis of therapist observed admission motor FIM™ score, cognitive FIM™ score, and age.

those in the skilled nursing facility (SNF) to each other and both to the gold standard. The following steps are applied to achieve case-mix adjustment.

1. Use **Table 1** to classify patients into their FRG based on one or more of the following bits of information: admission motor FIM™ score, admission cognitive FIM™ score, and age. The patient's motor FIM™ score is determined by summing observed 1–7 rated performances on the 13 motor FIM™ items and 5 cognitive items into separate subscales.
2. Use **Figure 1** to find the patient's FRG on the x axis. Find the patient's motor FIM™ discharge score on the y axis.
3. Mark above the patient's assigned DMF-FRG where the patient's motor FIM™ discharge status falls according to expected values on the y axis.
4. Determine if the patient's actual motor FIM™ score at discharge is below the 25 percentile (lower than the gold standard), between the 25th and 75th percentile (within the expected gold standard range), or above the 75th percentile (higher than the expected gold standard). The lower limit of the vertical bar represents the 25th and the upper limit the 75th percentile.
5. Count the number of patients over the CQI cycle (perhaps a quarter year) whose discharge motor FIM™ scores are below, within, and above the benchmarks.
6. Calculate the proportions of patients below, within, and above the ranges for their case-mix adjusted functional outcome benchmarks. Determine the proportions for the entire stroke population seen and stratified into persons being treated in the SNF and IRF. If the SNF and IRF are achieving comparable outcomes, the proportions of patients below, within, and above their DMF-FRG benchmarks would be comparable.

As noted previously, with constraints related to the prospective payer system (PPS), cost-containment efforts, and the shifting of care away from IRFs into less costly venues as illustrated by the vignette, the benchmarks represent

high targets. When this approach to case-mix–adjusted quality improvement is applied, it is important to recognize that comparison to the national high gold standard would be less important than internal comparisons over time within your health system. Repeat the analysis steps 1–6 through subsequent cycles of quality improvement. Compare the proportions of patients in your health care system below, within, and above their FRG-specific benchmark over time. The goal would be to develop clinical programs that will improve the quality and efficiency of care provided as evidenced by shifting increasing proportions of patients from the lower into the expected benchmark range and even higher relative to the high gold standard.

This approach is powerful but limited. Additional clinical factors such as degree of paralysis, visual impairments, aphasias, and comorbidity, while clinically important determinants of outcome, will in part be captured by patients' initial functioning. However, such factors may explain variation in outcomes over and above the FRGs. For other outcomes of interest, such as risks of nosocomial infection, other sets of factors will be important. Ultimately, multivariate regression models that consider the independent effects of many different risk factors on the outcome will lead to the best risk adjustment for obtaining expected values. The development of such models depends on careful review of the literature targeted to the particular outcome and the empirical derivation of the models applying large populations of patients.

Conclusions

Risk adjustment is an essential tool for monitoring and improving stroke rehabilitation care. Many aspects of risk adjustment are complex, but the basic principles can be used by clinicians and managers to ensure that achieved outcomes for stroke patients mirror those expected based on patient complexity. Appropriate case-mix adjustment could contribute to our capacity to measure interventions and program changes associated with meaningful improvements in outcomes for complex patients with diverse rehabilitation needs. The simple case-mix adjustment illustrated is appropriate to addressing functional outcomes. Functional outcomes are only one of many essential outcomes appropriate to CQI. Different risk-adjustment strategies would be necessary to address other types of outcomes or processes of care. Moreover, optimal approaches to CQI are generally multifactorial, addressing multiple indicators simultaneously and their associations as illustrated elsewhere in this article series.

Acknowledgments

This material is the result of work supported with resources and the use of facilities at the Edith Nourse Rogers Memorial Veterans Hospital, Bedford, Massachusetts. M. Stineman was supported in part through the Agency for Healthcare Research and Quality (AHRQ) Comparative Effectiveness of Alternative Levels of Stroke Rehabilitation grant 1- R01 HS018540-01.

REFERENCES

1. Iezzoni LI. Risk adjusting rehabilitation outcomes. *Am J Phys Med Rehabil.* 2004:83;316–326.
2. Baker DW, Einstadter D, Thomas CL, Husak SS, Gordon NH, Cebul RD. Mortality trends during a program that publically reported hospital performance. *Med Care.* 2002;40:879–890.
3. Iezzoni LI. Range of risk factors. In: Iezzoni LI, ed. *Risk Adjustment for Measuring Health Care Outcomes.* Chicago: Health Administration Press; 2003:33–70.
4. Berlowitz DR, Brandeis GH, Morris JN, et al. Deriving a risk-adjustment model for pressure ulcer development using the Minimum Data Set. *J Am Geriatr Soc.* 2001;49:866–871.
5. Berlowitz DR, Brandeis GH, Anderson JJ, et al. Evaluation of a risk-adjustment model for pressure ulcer development using the Minimum Data Set. *J Am Geriatr Soc.* 2001;49:872–876.
6. Rubin HR, Wu AW. The risk of adjustment. *Med Care.* 1992;30:973–975.
7. Fetter RB, Shin Youngsoo, Freeman JL, Averill RF, Thompson JD. Case mix definition by Diagnosis-Related Groups. *Med Care.* 1980; 18(suppl 2):1–53.

8. Fries BE. Schneider DP, Foley WJ, Gavazzi M, Burke R, Cornelius E. Refining a case-mix measure for nursing homes: Resource Utilization Groups (RUG-III). *Med Care.* 1994;32:668–685.
9. Knaus WA, Wagner DP, Zimmerman JE, Draper EA. Variations in mortality and length of stay in intensive care units. *Ann Intern Med.* 1993;118:753–761.
10. Charlson ME, Pompei EP, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis.* 1987;40:373–383.
11. Hoenig H, Sloane R, Horner RD, Zolkewitz M, Reker D. Differences in rehabilitation service s and outcomes among stroke patients cared for in veterans hospitals. *Health Serv Res.* 2001;35:1293–1315.
12. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. *Med Care.* 1998;36:8–27.
13. Weiner JP, Starfield BH, Steinwachs DM, Mumford LM. Development and application of a population-oriented measure of ambulatory care case-mix. *Med Care.* 1991;29:452–472.
14. Ellis RP, Pope GC, Iezzoni L, Ayanian JZ, Bates DW, Burstin H, et al. Diagnosis-based risk adjustment for Medicare capitation payments. *Health Care Financ Rev.* 1996;17:101–128.
15. Berlowitz DR, Hoenig H, Cowper DC, Duncan PW, Vogel B. Impact of comorbidities on stroke rehabilitation outcomes: Does the method matter? *Arch Phys Med Rehabil.* 2008;89:1903–1906.
16. Stineman MG, Tassoni CJ, Escarce JJ, et al. Development of function-related groups version 2.0: a classification system for medical rehabilitation. *Health Serv Res.* 1997;32:529–548.
17. Granger CV, Cotter AC, Hamilton BB, Fiedler RC. Functional assessment scales: a study of persons after stroke. *Arch Phys Med Rehabil.* 1993;74(2):133–138.
18. Pine M, Jones B, Lou Y. Laboratory values improve predictions of hospital mortality. *Int J Qual Healthcare.* 1998;10:491–501.
19. Davenport RJ, Dennis MS, Warlow CP. Effects of correcting outcome data for case mix: an example from stroke medicine. *BMJ.* 1996;312: 1505–1508.
20. Daley J, Khuri SF, Henderson W, et al. Risk adjustment of the postoperative morbidity rate for the comparative assessment of the quality of surgical care: results of the National Veterans Affairs Surgical Risk Study. *J Am Coll Surg.* 1997;185: 328–340.
21. Iezzoni LI, Ash AS, Shwartz M, Daley J, Hughes JS, Mackiernan YD. Judging hospitals by severity adjusted mortality rates: the influence of the severity-adjustment method. *Am J Public Health.* 1996;86:1379–1387.
22. Iezzoni LI, Shwartz M, Ash AS, Mackiernan YD. Predicting in-hospital mortality for stroke patients: Results differ across severity-measurement methods. *Med Decis Making.* 1996;16:348–356.
23. Bates BE, Stineman MG. Outcome indicators for stroke: application of an algorithm treatment across the continuum of postacute rehabilitation services. *Arch Phys Med Rehabil.* 2000;81: 1468–1478.
24. Bagg SD, Pombo AP, Hopman WM. Toward benchmarks for stroke rehabilitation in Ontario, Canada. *Am J Phys Med Rehabil.* 2006;85: 971–976.
25. Stineman MG, Granger CV. Outcome, efficiency, and time-trend pattern analyses for stroke rehabilitation. *Am J Phys Med Rehabil.* 1998;77: 193–201.
26. Stineman MG, Goin JE, Granger CV, Fiedler R, Williams SV. Discharge motor FIM-function related groups. *Arch Phys Med Rehabil.* 1997;78: 980–985.
27. Pronovost PJ, Berenholtz SM, Needham DM. Translating evidence into practice: a model for large scale knowledge translation. *BMJ.* 2008;337:a1714.

Factors Associated with Favorable Drinking Outcome 12 Months After Hospitalization in a Prospective Cohort Study of Inpatients with Unhealthy Alcohol Use

Nicolas Bertholet, MD, MSc[1,2,7], Debbie M. Cheng, ScD[1,3], Tibor P. Palfai, PhD[4,5], and Richard Saitz, MD, MPH[1,5,6]

[1]Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, Boston Medical Center and Boston University School of Medicine, Boston, MA, USA; [2]Alcohol Treatment Center, Centre Hospitalier Universitaire Vaudois and University of Lausanne, Lausanne, Switzerland; [3]Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA; [4]Department of Psychology, Boston University, Boston, MA, USA; [5]Youth Alcohol Prevention Center, Boston University School of Public Health, Boston, MA, USA; [6]Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA; [7]Clinical Epidemiology Center and Alcohol Treatment Center, CHUV, Lausanne, Switzerland.

BACKGROUND: Prevalence of unhealthy alcohol use among medical inpatients is high.
OBJECTIVE: To characterize the course and outcomes of unhealthy alcohol use, and factors associated with these outcomes.
DESIGN: Prospective cohort study.
PARTICIPANTS: A total of 287 medical inpatients with unhealthy alcohol use.
MAIN MEASURES: At baseline and 12 months later, consumption and alcohol-related consequences were assessed. The outcome of interest was a favorable drinking outcome at 12 months (abstinence or drinking "moderate" amounts without consequences). The independent variables evaluated included demographics, physical/sexual abuse, drug use, depressive symptoms, alcohol dependence, commitment to change (Taking Action), spending time with heavy-drinking friends and receipt of alcohol treatment (after hospitalization). Adjusted regression models were used to evaluate factors associated with a favorable outcome.
KEY RESULTS: Thirty-three percent had a favorable drinking outcome 1 year later. Not spending time with heavy-drinking friends [adjusted odds ratio (AOR) 2.14, 95% CI: 1.14–4.00] and receipt of alcohol treatment [AOR (95% CI): 2.16(1.20–3.87)] were associated with a favorable outcome. Compared to the first quartile (lowest level) of Taking Action, subjects in the second, third and highest quartiles had higher odds of a favorable outcome [AOR (95% CI): 3.65 (1.47, 9.02), 3.39 (1.38, 8.31) and 6.76 (2.74, 16.67)].
CONCLUSIONS: Although most medical inpatients with unhealthy alcohol use continue drinking at-risk amounts and/or have alcohol-related consequences, one third are abstinent or drink "moderate" amounts without consequences 1 year later. Not spending time with heavy-drinking friends, receipt of alcohol treatment and commitment to change are associated with this favorable outcome. This can inform efforts to address unhealthy alcohol use among patients who often do not seek specialty treatment.

KEY WORDS: unhealthy alcohol use; medical inpatients; factors associated with drinking and consequences.
J Gen Intern Med 25(10):1024–9
DOI: 10.1007/s11606-010-1382-1

Received September 25, 2009
Revised February 8, 2010
Accepted March 29, 2010
Published online May 18, 2010

INTRODUCTION

Unhealthy alcohol use (alcohol consumption that increases the risk of health consequences and includes abuse and dependence) is a major public health concern[1,2]. In primary care settings the prevalence of unhealthy alcohol use is 7 to 20% or more, with most people not suffering from alcohol dependence[3]. However, in medical hospital settings, the proportion of patients with unhealthy alcohol use who meet the criteria for alcohol dependence is high[4]. For example, in four general hospitals in Germany, Freyer-Adam et al. found that 61% of inpatients with unhealthy alcohol use had alcohol dependence[5]. In a large urban safety-net hospital in the US (the sample for the current study), the proportion was 77%[4,6]. As such, the problem of unhealthy alcohol use in inpatient medical settings is likely to differ from that in other (particularly outpatient) health care settings.

In general populations, the natural history of drinking among those with dependence has been well studied, and social and personal factors have been identified as predictors of natural recovery. Epidemiologic studies indicate that there is a substantial proportion of individuals with alcohol dependence who will be in recovery 12 months later[7]. Age and participation in self-help or treatment affect the course of substance dependence and male gender, depression, heroin and cocaine use, divorce and low level of education are related to worse outcome[8]. "Resolution" (abstinence for more than 2 years) has been linked to heavier drinking practices and negative life events during the year before the onset of abstinence[9].

Nevertheless, the course of drinking and predictors of favorable drinking outcome among medical inpatients are not

well known. Describing both consumption and consequence outcomes allows assessment of a range of outcomes among diverse patients. It is of clinical interest among inpatients, especially because some may choose a "moderate" drinking goal during brief counseling sessions, and may do so without negative consequences even if previously diagnosed with dependence. Moderate drinking may be an appropriate goal if patients can do so without consequences.

Understanding the course of unhealthy alcohol use and predictors of favorable consumption and consequence outcomes may help clinicians to tailor advice and treatment planning, and to develop interventions for medical inpatients. The latter is important given the lack of robust evidence for the efficacy of brief interventions in this setting[5,10,11].

Therefore, we studied a prospective cohort of medical inpatients with unhealthy alcohol use to determine the course of alcohol use and consequences, and factors associated with favorable drinking outcomes, 1 year after hospitalization. We hypothesized that factors such as male gender, low socioeconomic status, depression, physical or sexual abuse, illegal drug use, presence of alcohol dependence and social pressure to drink would be associated with unfavorable outcome, and factors such as readiness to change and receipt of specialized alcohol treatment (including self-help) would be associated with favorable outcome.

METHODS

Data were collected by interview with medical inpatients at an urban academic hospital who were drinking risky amounts of alcohol [>14 standard drinks (14 g of pure alcohol)/week or ≥5 drinks on an occasion for men, >11 drinks/week or ≥4 drinks on an occasion for women and persons aged over 65 years]. This cohort was prospectively followed for 1 year. Subjects were participants in a randomized trial of a single brief motivational interviewing counseling session (compared with no brief motivational counseling); the intervention had no significant effect on drinking or alcohol consequences[6].

Research associates approached all patients aged ≥18 whose physicians did not decline the contact. Individuals fluent in English or Spanish who gave consent were asked to complete a screening interview. Eligibility criteria were: currently drinking risky amounts (as above), two contacts to assist follow-up, no plans to move from the area for the next year and a Mini-Mental State Examination score of ≥21. During the screening interview, subjects completed a 1–10 visual analog scale for readiness to change ("How ready are you to change your drinking habits"). Subjects who refused participation were more likely to be Black (45% vs 31%) and to drink greater amounts of alcohol (median 24 vs 18 drinks per week) compared to eligible subjects who enrolled, but were similar regarding readiness to change measured on a 1–10 visual analog scale.

At study entry we assessed demographics, principal admitting and alcohol-attributable medical diagnoses (by medical record review), alcohol use disorder diagnosis [assessed using the Composite International Diagnostic Interview (CIDI) Alcohol Module][12,13], education, homelessness, heroin use, cocaine use, physical or sexual abuse before the age of 18, and whether or not the subject spent time with heavy or problem drinkers (reflecting social pressure to drink). Not spending time with heavy-drinking friends was assessed with the question "How many of the people you spend time with are heavy or problem drinkers?" and later dichotomized into none vs. any. Baseline measures of health-related quality-of-life (QOL) [Short-Form Health Survey, Physical Component Summary (PCS) score][14], depressive symptoms, and readiness to change alcohol use [problem recognition and commitment to change drinking with the "Perception of Problem" (range 10–50) and "Taking Action" (range 6–30) scales, respectively] were also used. These latter two scales were determined based on a factor analysis of the Stages of Change Readiness Treatment and Eagerness Scale (SOCRATES) in this sample[15]. The Taking Action scale questions assess both actions to facilitate change that already occurred and commitment to change, with a higher score indicating a higher level of having taken action and commitment to change[16]. At 12 months, receipt of treatment since study entry was assessed by self-report [hospital detoxification, any treatment for alcohol problems (including counseling or therapy), Alcoholics Anonymous meetings, self-help, mutual help or other 12-step programs for alcohol problems or medication prescribed by a physician to prevent them from drinking]. At study entry and 12 months later, we assessed alcohol consumption with a validated 30-day calendar method (Timeline Followback)[17] and alcohol-related consequences with the Short Inventory of Problems (SIP)[18]. The outcome of interest was favorable drinking at 12 months, defined as abstinence or drinking "moderate" amounts [i.e., less than at-risk amounts, defined above except >7 (not 11) drinks per week was the cutoff for women and the elderly] without consequences. The outcome definition was based on a procedure described and validated by Cisler and Zweben[19–21] that classified drinkers according to two factors: whether they drank at-risk amounts and whether they experienced alcohol-related consequences. This composite outcome index was created to capture a broader range of clinically relevant outcomes, knowing that patients with unhealthy alcohol use may choose to keep on drinking but at lower levels and without suffering from alcohol-related consequences.

The following factors were tested: education, marital status, homelessness, physical or sexual abuse before the age of 18, heroin or cocaine use, elevated depressive symptoms, presence of alcohol dependence, readiness to change measure, spending time with heavy-drinking friends and receipt of alcohol treatment after hospitalization (evaluated at the 12-month assessment).

Confounders were defined a priori based on literature and clinical experience, and included: age, gender, race/ethnicity, randomization group, PCS and drinking (drinks per day, past 30 days) at study entry. We included a measure of drinking at study entry as a confounder because alcohol consumption level is known to be one of the strongest predictors of subsequent drinking.

Analyses

We used an iterative model-building procedure to identify factors associated with favorable drinking. Each factor of interest was entered in a separate model adjusted for potential confounders (i.e., in "minimally adjusted models"): age, gender, race/ethnicity, randomization group, physical health-related QOL (PCS score), presence of an alcohol-attributable principal diagnosis at hospital admission and drinks per day at study entry. The potential confounders were selected a priori based on literature and clinical experience.

Prior to regression modeling, we assessed bivariate correlations between all independent variables and covariates. To

avoid potential collinearity, no pair of variables with Spearman correlation coefficient >0.40 was included in the same model. Because it was correlated with other factors of interest (drinks per day, past 30 days, r=0.54; alcohol dependence, r=0.54; receipt of alcohol treatment, r=0.44; depressive symptoms, r=0.44), the Perception of Problem (PP) scale was excluded from further analyses. Since other work suggests that the PP scale and other variables reflecting perception of alcohol problems are markers of severity and since the PP scale was correlated with other better markers of severity, we excluded it from further multivariable analyses. We nevertheless report unadjusted models for PP, since measures of perception of alcohol problem are often used in the literature. No other pairs were correlated >0.40. In an unadjusted logistic regression model, subjects in the highest quartile (highest level) of PP had 2.21 (95% CI: 1.07, 4.56) times the odds of an *unfavorable* drinking outcome compared to the lowest quartile.

Factors significantly associated with the drinking outcome at an alpha level of 0.05 in these "minimally adjusted models" were included together in a single multivariable model along with confounders. Factors that were no longer significant at an alpha level of 0.05 in the multivariable model were removed one at a time to obtain the final model.

All analyses were adjusted for randomization group (i.e., assignment to brief intervention) at baseline. Analyses were performed using SAS software (version 9.1; SAS Institute, Cary, NC).

RESULTS

Of the 5,813 patients screened, 986 were drinking risky amounts. Of these, 462 were not eligible for entry into the cohort, and 183 were eligible but declined. Of the 341 subjects who were eligible and consented to be in the cohort, 287 (84%) had complete data at 12 months and were included in these analyses. The baseline characteristics of the 287 subjects are presented in Table 1. The five most prevalent principal diagnoses at hospital admission were: rule out myocardial infarction (n=50), pancreatitis (n=31), cellulitis (n=20), asthma (n=19) and pneumonia (n=19). Subjects who completed the 12-month follow-up did *not* differ significantly (alpha level 0.05) from those lost to follow-up with respect to the baseline characteristics presented in Table 1.

At 12 months, most subjects (63%) were drinking risky amounts, 29% were abstinent, and a few were drinking moderate amounts (with or without consequences) (8%). At 12 months, 33% had a favorable drinking outcome [i.e., they were abstinent (29%) or drinking moderate amounts without consequences (4%)] (Fig. 1).

Table 2 presents unadjusted logistic regression models for all factors of interest and confounders, and the final model developed from the iterative model building procedure. Elevated depressive symptoms, Taking Action, not spending time with heavy-drinking friends and receipt of alcohol treatment after hospital discharge were associated with a favorable drinking outcome in both unadjusted and "minimally adjusted models." These four variables were entered simultaneously with the a priori defined potential confounders in an adjusted logistic regression model. The depressive symptom variable was no longer significant in the multivariable model (p=0.2) and was therefore excluded from the final model. In the final model (Table 2), compared to the first quartile (lowest level) of Taking Action, subjects in the second, third and highest quartile had 3.65 (95% confidence interval, 1.47, 9.02), 3.39 (95% CI 1.38, 8.31) and 6.76 (95% CI 2.74, 16.67) times the odds of a favorable drinking outcome, respectively. Not spending time with heavy-drinking friends [adjusted odds ratio (AOR) 2.14; 95% CI 1.14, 4.00] and receipt of alcohol treatment after hospital discharge during the past year (from baseline to 12-month assessment; AOR 2.16, 95% CI 1.20, 3.87) were associated with a favorable drinking outcome. The Hosmer and Lemeshow chi-square test suggested acceptable model fit of the final logistic regression model (p=0.73).

Table 1. Characteristics at Study Entry of 287 Medical Inpatients Identified by Screening with Unhealthy Alcohol Use

Characteristics	
Age, mean (SD)	44.4 (10.6)
Female, n (%)	86 (30)
Currently married, n (%)	32 (11.2)
Education (years), mean (SD)	11.9 (2.5)
Homelessness, n (%)	73 (25)
Race/ethnicity	
African-American, n (%)	133 (46)
White, n (%)	108 (38)
Hispanic, n (%)	24 (8)
Other, n (%)	22 (8)
DSM IV Alcohol Diagnosis	
Alcohol dependence, n (%)	223 (78)
Alcohol abuse, n (%)	13 (4)
No alcohol use disorder diagnosis, n (%)	51 (18)
Alcohol consumption (drinks per day, past 30 days), mean (SD)	6.8 (8.9)
SOCRATES	
Perception of Problem, mean (SD)	35.5 (11.1)
Taking Action, mean (SD)	21.2 (5.8)
Depressive symptoms (CES-D score ≥16), n (%)	203 (71)
Heroin or cocaine use (past 30 days), n (%)	74 (26)
Physical abuse before age 18, n (%)	116 (41)
Sexual abuse before age 18, n (%)	66 (23)
Not spending time with heavy-drinking friends (less social pressure to drink), n (%)	84 (29)
Alcohol-attributable principal diagnosis at hospital admission, n (%)	42 (15)
Receipt of alcohol treatment including self-help	127 (44.6)

Homelessness was defined as more than 1 night spent on the streets or in a shelter over the past 3 months
SOCRATES: Stages of Change Readiness Treatment and Eagerness Scale
DSM IV: Diagnostic and Statistical Manual of Mental Disorders, 4th edn
CES-D: Center for Epidemiologic Studies Depression scale
Alcohol-attributable diagnosis includes any of the following: acute alcoholic cirrhosis of the liver, alcoholic cardiomyopathy, alcoholic gastritis, alcoholic hepatits, alcohol intoxication, alcoholic liver damage, alcoholic fatty liver, alcoholic pellagra, alcoholic polyneurpoathy, alcoholic withdrawal, alcoholic withdrawal convulsion, alcoholic withdrawal delirium, alcoholic withdrawal hallucinosis, other alcoholic psychosis, alcoholic amnestic syndrome, other alcoholic dementia, alcoholic pancreatitis or other diagnoses considered alcohol-attributable by the investigator (e.g., holiday heart, alcoholic ketoacidosis, alcohol-related rhabdomyolisis)

DISCUSSION

We investigated unhealthy alcohol use outcomes and factors associated with a favorable drinking outcome (abstinence or moderate drinking without consequences) at 12 months in opportunistically screened medical inpatients who were not

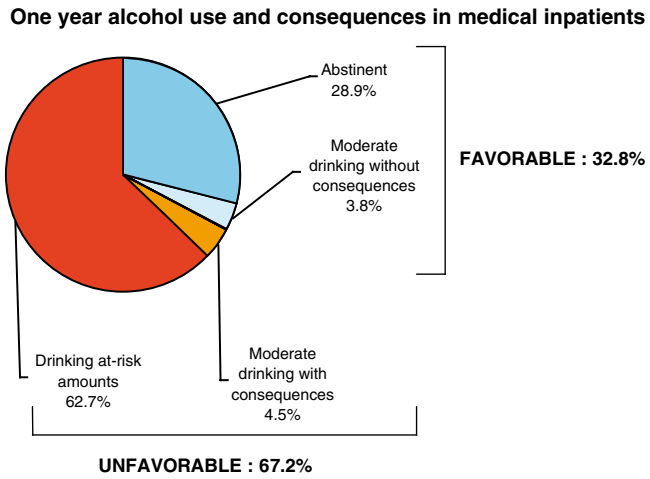


Figure 1. Note: Alcohol consumption was assessed with the 30-day Timeline Followback method. Alcohol consequences were assessed with the Short Inventory of Problems (SIP). The SIP is a 15-item questionnaire that assesses, over the past 3 months, the presence of alcohol-related consequences in various dimensions: physical, interpersonal, intrapersonal, impulse control and social responsibility.

necessarily seeking specialty alcohol treatment. Most continued to drink amounts that risk health consequences and/or have such consequences, but one third had a favorable drinking outcome. Abstinence was the most likely favorable drinking outcome. Few were drinking moderate amounts, with or without consequences.

In 1976, Imber et al. studied the natural history of drinking in male general hospital inpatients with alcohol dependence; 19% were abstinent 1 year later[22]. More than 30 years later, we found a similar though higher proportion of abstinence. The more favorable course might be due to a sample that consisted not only of subjects with alcohol dependence (though absence of alcohol dependence was not predictive of favorable drinking outcome in our sample). In the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC), Dawson et al. showed that among individuals with alcohol dependence, a year later, 17.7% were drinking "low-risk" amounts, and 18.2% were abstainers; approximately half of those with a favorable consumption outcome were still drinking[7]. Similarly, Sobell et al., in two Canadian general population studies of individuals with alcohol dependence, found that low-risk drinking accounted for 40 and 60% of all cases of recovery[7,23]. Yet in our study, even though some subjects did not have alcohol dependence, favorable drinking outcome most often consisted of abstinence. Although reasons for the different observations are not clear, the setting and severity of the general hospital sample likely account for them in part.

In primary care, where the prevalence of alcohol dependence is lower, a similar proportion of screen-identified patients was not drinking risky amounts 6 months later[24]. Similarly, in a study of screening and brief intervention conducted among inpatients with unhealthy alcohol use (without dependence), 46% of the controls did not report any alcohol problems 12 months later. On average, they decreased their daily alcohol consumption by 24 g of alcohol (from 70 g). In most brief intervention studies, a substantial decrease in drinking has been observed in the groups that did not receive any interven-

Table 2. Associations with Favorable Drinking Outcome 1 Year After General Medical Hospitalization of 287 Patients with Unhealthy Alcohol Use: Unadjusted and Final Logistic Regression Models

	Unadjusted model		Final model	
	OR	95% CI	AOR	95% CI
Factors of interest				
Education (for a 1 year difference)	1.100	0.992, 1.219		
Currently married	0.651	0.280, 1.511		
Homelessness	1.475	0.843, 2.582		
Physical or sexual abuse before age 18	0.981	0.595, 1.616		
Heroin or cocaine use	1.088	0.619, 1.914		
Depressive symptoms (CES-D ≥ 16)	2.224	1.211, 4.081		
Alcohol dependence	1.294	0.700, 2.392		
Taking Action (lowest quartile = reference group)				
2nd quartile	3.362	1.425, 7.930	3.645	1.473, 9.017
3rd quartile	3.222	1.376, 7.542	3.386	1.380, 8.308
Highest (4th) quartile	6.443	2.760, 15.043	6.758	2.740, 16.667
Not spending time with heavy-drinking friends	1.896	1.109, 3.241	2.137	1.142, 4.000
Receipt of alcohol treatment in the past 12 months	1.959	1.183, 3.247	2.160	1.204, 3.874
Possible confounders				
Age (for a 1-year difference)	1.003	0.980, 1.027	0.992	0.964, 1.022
Gender (female)	1.113	0.648, 1.911	1.043	0.569, 1.912
Race/ethnicity (non-white vs white)	0.842	0.505, 1.406	0.676	0.375, 1.221
Randomization group (intervention)	0.977	0.593, 1.610	0.927	0.529, 1.625
PCS	0.987	0.959, 1.015	0.975	0.944, 1.007
Drinking at baseline (drinks per day)	1.008	0.981, 1.035	1.003	0.973, 1.035
Alcohol-attributable principal diagnosis at hospital admission	1.727	0.857, 3.479	2.153	1.005, 4.610

OR: odds ratio
AOR: adjusted odds ratio
CI: confidence interval
Taking Action: subscale of the Stages of Change Readiness Treatment and Eagerness Scale
CES-D: Center for Epidemiologic Studies Depression Scale
PCS: Physical component summary

tion[10,25–27]. The course of unhealthy alcohol use tends to involve self-change with or without formal help, with a substantial proportion of individuals either abstaining or drinking moderate amounts without consequences a year later. These changes take place for individuals in the general population and for individuals that have contacts with the health care system. These changes may be due to self-change, life events and experiences, notably the accumulation of negative events, as well as assessment effects—all things that happen regardless of interventions.

Few studies have investigated the course and factors associated with outcomes of unhealthy alcohol use in medical inpatients. Although people who enroll in trials differ from those who do not, and although assessment effects can affect outcomes, cohorts of subjects from randomized controlled trials can provide some relevant information. In addition, we identified factors associated with favorable drinking outcome: Taking Action (a measure of actions towards facilitation of change and commitment to change, which can be considered a specific subcategory of readiness to change), not spending time with heavy-drinking friends (which can be seen as a proxy measure for the social pressure to drink) and receipt of alcohol treatment over the past 12 months were positively associated with a favorable drinking outcome. Even though usually considered markers of severity or predictors of poor outcome, and contrary to our hypotheses, a diagnosis of alcohol dependence, drug use, low education level and homelessness were not associated with drinking outcome[8,28].

These results suggest the importance of commitment to change and action toward facilitation of change as valuable targets for interventions[29]. As shown in other studies, individuals who tend to have some intention or commitment to reduce their drinking when seen in a hospital will have a better prognosis[30]. Self-report of receipt of alcohol treatment between study entry and 12 months later was associated with favorable drinking outcome. This supports the current knowledge on treatment efficacy[2,31,32]. The observed positive predictive effect of not having heavy-drinking friends on favorable drinking outcome is also consistent with studies indicating the negative impact of the social environment on drinking, notably the impact of social pressure to drink and its negative impact on relapse risk[33]. Our results add to the evidence that the absence of a heavy-drinking social environment is associated with a better drinking prognosis for individuals with unhealthy alcohol use. Future research may explore relationships between alcohol use behaviors and social networks in order to determine if the same social network effects found in tobacco cessation can be identified for alcohol use[34].

The fact that factors usually considered predictors of poor outcome in outpatients (i.e., diagnosis of alcohol dependence, drug use, low education level, homelessness, childhood physical or sexual abuse) were not associated with drinking outcomes in our study is of interest. This may have been due to a lack of power, or alternatively to intrinsic differences in hospitalized patients and ambulatory patients with unhealthy alcohol use. Specifically, our study of hospitalized patients may have examined a more homogeneous and sicker population than usually enrolled in general population studies. Individuals with less severe social and health problems tend to access the health care system less and were therefore less likely to be included in our study. Notably, the present study showed a high prevalence of alcohol dependence among individuals with unhealthy alcohol use (i.e., most patients who screened positive had alcohol dependence). Even though the prevalence of dependence is usually high in screen-positive hospitalized patients, the fact that the study population was recruited at an urban safety net hospital may explain an even higher prevalence. Nevertheless, our findings suggest that these poor prognostic factors should not be seen as insurmountable obstacles when addressing unhealthy alcohol use among medical inpatients.

Study limitations should be considered when interpreting our findings. First, we evaluated the course of unhealthy alcohol use in a sample of subjects that was included in a randomized trial. It is unlikely that the intervention affected our results since we controlled for it in analyses and the trial had negative results. The subjects agreed to participate in a study in which they could receive alcohol counseling. This might have resulted in a selection of individuals more prone to behavior change or more motivated to change; however, it might also have led to selection of patients who were interested in counseling because they thought they could not change without it. Subjects who refused participation were similar to subjects who participated regarding readiness to change completed during the screening interview. Study subjects may also have had courses not representative of natural history due to assessment effects. If this is the case, then the course of unhealthy alcohol use in medical inpatients would be even worse than what we observed. Second, our study was able to identify associations over time but was not designed to study causation. This study is a secondary observational analysis, thus observed associations may not be causal and analyses may be underpowered. The latter may explain why some factors were not significantly associated with drinking outcome, though despite this possibility, other potentially useful and easily assessable clinical factors were associated with outcome. Third we grouped together various treatment modalities and are therefore not able to distinguish between them, though all are known to have efficacy[31]. The present cohort was followed for 12 months. This could be seen as short with regard to drinking outcomes. Our results should be replicated in cohorts with longer follow-up and with multiple assessment time points. Nevertheless, the literature suggests that 12-month outcomes are indicative of longer term functioning[35,36].

This study has notable strengths. We used a large prospective sample identified by screening patients in a general health care setting, with a high follow-up rate. Prospective observational studies are the ideal approach to studying the outcomes and their predictors. Our subjects were well-characterized using validated assessments. We also used a composite outcome of drinking and consequences that has been validated and that has clinical significance[20,21].

Our results bring additional information to clinicians treating medical inpatients, a population where unhealthy alcohol use is common. In this setting, one third of the patients will be abstinent a year later. Thus, some optimism regarding the natural history of alcohol use in this population is reasonable. Our results also suggest that homelessness, drug use and depressive symptoms, usually considered markers of severity or predictors of poor outcome, may not have a large negative impact on drinking outcome in medical inpatients. The presence of these markers should not prevent clinicians from addressing unhealthy alcohol use and should not lead them

(or their patients) to have a pessimistic view of the drinking prognosis.

The factors identified as being associated with favorable outcome could be useful for clinicians, since they are potentially amenable to change with clinician assistance. Clinicians should therefore be encouraged to target commitment to change in their discussions with patients and help them to take actions towards change. Similarly, linking medical inpatients with unhealthy alcohol use to alcohol treatment and encouraging changes in their social environment to decrease the social pressure to drink may increase the likelihood of a favorable outcome. Since abstinence was the most likely favorable outcome, clinicians should suggest to patients in this setting that abstinence should remain the preferred treatment goal. Clinicians should also keep in mind that factors usually seen as predictive of poor outcomes may not be obstacles among medical inpatients to the degree they may be in other populations of patients with unhealthy alcohol use.

Sources of funding: *Nicolas Bertholet was supported by the Swiss National Science Foundation and Fondation Suisse de Recherche sur l'Alcool. The randomized trial was supported by the National Institute on Alcohol Abuse and Alcoholism (a grant to Boston Medical Center, Dr. Richard Saitz, Principal Investigator, grant R01 AA 12617).*

Conflict of interest: *None disclosed*

Corresponding Author: *Nicolas Bertholet, MD, MSc; Clinical Epidemiology Center and Alcohol Treatment Center, CHUV, Mont Paisible 16 1011, Lausanne, Switzerland (e-mail: Nicolas.Bertholet@chuv.ch).*

REFERENCES

1. **Rehm J, Room R, Monteiro M, et al.** Alcohol as a risk factor for global burden of disease. Eur Addict Res. 2003;9:157–64.
2. **Saitz R.** Clinical practice. Unhealthy alcohol use. N Engl J Med. 2005;352:596–607.
3. **Institute of Medicine.** Broadening the base of treatment for alcohol for alcohol problem. Washington, DC: National Academy Press; 1990.
4. **Saitz R, Freedner N, Palfai TP, Horton NJ, Samet JH.** The severity of unhealthy alcohol use in hospitalized medical patients. The spectrum is narrow. J Gen Intern Med. 2006;21:381–5.
5. **Freyer-Adam J, Coder B, Baumeister SE, et al.** Brief alcohol intervention for general hospital inpatients: A randomized controlled trial. Drug Alcohol Depend. 2008;93:233–43.
6. **Saitz R, Palfai TP, Cheng DM, et al.** Brief intervention for medical inpatients with unhealthy alcohol use: a randomized, controlled trial. Ann Intern Med. 2007;146:167–76.
7. **Dawson DA, Grant BF, Stinson FS, Chou PS, Huang B, Ruan WJ.** Recovery from DSM-IV alcohol dependence: United States, 2001–2002. Addiction. 2005;100:281–92.
8. **Moss HB, Chen CM, Yi HY.** Subtypes of alcohol dependence in a nationally representative sample. Drug Alcohol Depend. 2007;Dec 1;91 (2–3):149–58
9. **Tucker JA, Vuchinich RE, Rippens PD.** Different variables are associated with help-seeking patterns and long-term outcomes among problem drinkers. Addict Behav. 2004;29:433–9.
10. **Daeppen JB, Gaume J, Bady P, et al.** Brief alcohol intervention and alcohol assessment do not influence alcohol use in injured patients treated in the emergency department: a randomized controlled clinical trial. Addiction. 2007;102:1224–33.
11. **Emmen MJ, Schippers GM, Bleijenberg G, Wollersheim H.** Effectiveness of opportunistic brief interventions for problem drinking in a general hospital setting: systematic review. BMJ. 2004;328:318.
12. **Robins LN, Wing J, Wittchen HU, et al.** 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.
13. WHO. Composite International Diagnostic Interview (CIDI)(Core Version 2.0). Geneva, Swtizerland: World Health Organization; 1996.
14. **Ware JE Jr, Sherbourne CD.** Thirty-six item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care. 1992;30:473–83.
15. **Bertholet N, Dukes K, Horton NJ, Palfai TP, Pedley A, Saitz R.** Factor structure of the SOCRATES questionnaire in hospitalized medical patients. Addict Behav. 2009;34:568–72.
16. **Bertholet N, Cheng DM, Palfai TP, Samet JH, Saitz R.** Does readiness to change predict subsequent alcohol consumption in medical inpatients with unhealthy alcohol use? Addict Behav. 2009;34:636–40.
17. **Sobell LC, Sobell MB.** Alcohol Timeline Followback (TLFB) Users' Manual. Toronto, Canada: Addiction Research Foundation; 1995.
18. **Miller WR, Tonigan J, Longabaugh R.** The Drinker Inventory of Consequences (DrInC). An Instrument for Assessing Adverse Consequences of Alcohol Abuse. Test Manual. 4 vol. Bethesda: NIH; 1995.
19. **Zweben A, Cisler R.** Composite outcome measures in alcoholism treatment research: problems and potentialities. Subst Use Misuse. 1996;31:1783–805.
20. **Zweben A, Cisler RA.** Clinical and methodological utility of a composite outcome measure for alcohol treatment research. Alcohol Clin Exp Res. 2003;27:1680–5.
21. **Cisler RA, Zweben A.** Development of a composite measure for assessing alcohol treatment outcome: operationalization and validation. Alcohol Clin Exp Res. 1999;23:263–71.
22. **Imber S, Schultz E, Funderburk F, Allen R, Flamer R.** The fate of the untreated alcoholic. Toward a natural history of the disorder. J Nerv Ment Dis. 1976;162:238–47.
23. **Sobell LC, Cunningham JA, Sobell MB.** Recovery from alcohol problems with and without treatment: prevalence in two population surveys. Am J Public Health. 1996;86:966–72.
24. **Bertholet N, Horton NJ, Saitz R.** Changes in readiness and drinking in primary care patients with unhealthy alcohol use. Alcoholism Clin Exp Res. 2007;Blackwell.
25. **Bertholet N, Daeppen JB, Wietlisbach V, Fleming M, Burnand B.** Reduction of alcohol consumption by brief alcohol intervention in primary care: systematic review and meta-analysis. Arch Intern Med. 2005;165:986–95.
26. **Kaner EF, Beyer F, Dickinson HO et al.** Effectiveness of brief alcohol interventions in primary care populations. Cochrane Database Syst Rev. 2007:CD004148.
27. **D'Onofrio G, Pantalon MV, Degutis LC, et al.** Brief intervention for hazardous and harmful drinkers in the emergency department. Ann Emerg Med. 2008;51:742–750. e2.
28. **Gilman SE, Breslau J, Conron KJ, Koenen KC, Subramanian SV, Zaslavsky AM.** Education and race-ethnicity differences in the lifetime risk of alcohol dependence. J Epidemiol Community Health. 2008;62:224–30.
29. **Amrhein PC, Miller WR, Yahne CE, Palmer M, Fulcher L.** Client commitment language during motivational interviewing predicts drug use outcomes. J Consult Clin Psychol. 2003;71:862–78.
30. **Daeppen JB, Bertholet N, Gmel G, Gaume J.** Communication during brief intervention, intention to change, and outcome. Subst Abus. 2007;28:43–51.
31. **Miller WR, Wilbourne PL.** Mesa Grande: a methodological analysis of clinical trials of treatments for alcohol use disorders. Addiction. 2002;97:265–77.
32. **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;295:2003–17.
33. **Zywiak WH, Stout RL, Longabaugh R, Dyck I, Connors GJ, Maisto SA.** Relapse-onset factors in Project MATCH: the Relapse Questionnaire. J Subst Abuse Treat. 2006;31:341–5.
34. **Christakis NA, Fowler JH.** The collective dynamics of smoking in a large social network. N Engl J Med. 2008;358:2249–58.
35. **Maisto SA, Clifford PR, Stout RL, Davis CM.** Drinking in the year after treatment as a predictor of 3-year outcomes. J Stud Alcohol. 2006;67:823–32.
36. **Maisto SA, Clifford PR, Stout RL, Davis CM.** Moderate drinking in the first year after treatment as a predictor of 3-year outcomes. J Stud Alcohol Drugs. 2007;68:419–27.

Drug and Alcohol Dependence 112 (2010) 160–163

Contents lists available at ScienceDirect

Drug and Alcohol Dependence

journal homepage: www.elsevier.com/locate/drugalcdep

Short communication

Alcohol consumption patterns in HIV-infected adults with alcohol problems

Nicolas Bertholet[a,b,c,*], Debbie M. Cheng[a,d], Jeffrey H. Samet[a,e], Emily Quinn[f], Richard Saitz[a,g]

[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, 801 Massachusetts Avenue, 2nd Floor, Boston, MA 02118-2605, United States
[b] Clinical Epidemiology Center, Institute of Social and Preventive Medicine, Centre Hospitalier Universitaire Vaudois and University of Lausanne, Bugnon 17, Lausanne 1005, Switzerland
[c] Alcohol Treatment Center, Centre Hospitalier Universitaire Vaudois and University of Lausanne, Mont-Paisible 16, Lausanne 1005, Switzerland
[d] Department of Biostatistics, Boston University School of Public Health, 801 Massachusetts Avenue, 3rd Floor, Boston, MA 02118-2605, United States
[e] Department of Social and Behavioral Sciences, Boston University School of Public Health, 801 Massachusetts Avenue, 4th Floor, Boston, MA 02118-2605, United States
[f] DM-STAT Inc., 1 Salem Street Malden, MA 02148-5236, United States
[g] Youth Alcohol Prevention Center, Boston University School of Public Health, 580 Harrison Avenue, 3rd Floor, Boston, MA 02118-2605, United States

ARTICLE INFO

Article history:
Received 23 October 2009
Received in revised form 16 May 2010
Accepted 25 May 2010
Available online 25 June 2010

Keywords:
Alcohol
HIV
Drinking patterns

ABSTRACT

Objective: To understand patterns of alcohol consumption and baseline factors associated with favorable drinking patterns among HIV-infected patients.
Methods: We studied drinking patterns among HIV-infected patients with current or past alcohol problems. We assessed drinking status in 6-month intervals. Based on National Institute on Alcohol Abuse and Alcoholism guidelines a favorable drinking pattern was defined as not drinking risky amounts at each assessment or decreased drinking over time. All other patterns were defined as unfavorable. Logistic regression models were used to identify baseline factors associated with a favorable pattern.
Results: Among 358 subjects, 54% had a favorable drinking pattern with 44% not drinking risky amounts at every assessment, and 11% decreasing consumption over time. Of the 46% with an unfavorable pattern, 4% drank risky amounts each time, 5% increased, and 37% both decreased and increased consumption over time. Current alcohol dependence and recent marijuana use were negatively associated with a favorable pattern, while older age and female gender, and having a primary HIV risk factor of injection drug use were positively associated with a favorable pattern.
Conclusion: Many HIV-infected adults with alcohol problems have favorable drinking patterns over time, and alcohol consumption patterns are not necessarily constant. Identifying HIV-infected adults with a pattern of risky drinking may require repeated assessments of alcohol consumption.

1. Introduction

Alcohol use is common among human immunodeficiency virus (HIV)-infected adults and has negative health consequences. Over a third of HIV-infected veterans drink amounts associated with health consequences (Samet et al., 2007a,b). In another study, almost half of HIV-infected patients initiating primary care reported a high probability of having an alcohol use disorder (had two of more positive answers to the CAGE screening tool) (Mayfield et al., 1974; Samet et al., 2004b). Unhealthy alcohol use (the spectrum from drinking risky amounts through alcohol dependence) is more prevalent among HIV-infected patients than it is in the general population (Lefevre et al., 1995; Conigliaro et al., 2003; Samet et al., 2004b; Chander et al., 2006). Drug and unhealthy alcohol use have been linked to HIV-disease progression, HIV risk behaviors, and decreased adherence to antiretroviral therapy (Samet et al., 2007a,b). In adults with alcohol use disorders, social and personal factors (e.g., male gender, major depression, heroin use, cocaine use, divorce, and less education) have been linked to a worse prognosis (Moss et al., 2007). The environment has an impact on the course of drinking too (e.g., social pressure to drink has been linked with relapse among individuals with alcohol dependence) (Zywiak et al., 2006a,b).

There is sufficient evidence to date (though not specifically among those with HIV infection) to conclude that brief counseling in the primary care setting for nondependent unhealthy alcohol use can lead to a decrease in drinking (Bertholet et al., 2005; Kaner et al., 2007), including among injection drug users (Stein et al., 2002). Nevertheless, alcohol use disorders are often chronic conditions characterized by recurrent episodes, and few prospective studies have explored alcohol consumption over time. Since the treatment of HIV infection requires longitudinal care, it gives clinicians repeated opportunities to address alcohol use. A greater understanding of consumption over time in HIV-infected individuals with alcohol use disorders could help clinicians and researchers

* Corresponding author at: Alcohol Treatment Center, Mont-Paisible 16, 1011 Lausanne, CHUV, Switzerland. Tel.: +41 21 314 73 51; fax: +41 21 314 05 62.
E-mail address: nberthol@bu.edu (N. Bertholet).

0376-8716/$ – see front matter © 2010 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.drugalcdep.2010.05.009

better address these problems. Therefore, in a prospective cohort of HIV-infected adults with current or past alcohol problems (HIV-Longitudinal Interrelationships of Viruses and Ethanol [HIV-LIVE]), we studied patterns of alcohol consumption and factors associated with those patterns.

2. Methods

Participants were recruited between August 2001 and July 2003 with follow-up every 6 months through 2005. Recruitment occurred from a previous cohort study (HIV-Alcohol Longitudinal Cohort), an intake clinic for HIV-infected patients, HIV primary care and specialty clinics, a homeless respite facility, a methadone program, study flyers, and referrals from physicians, other participants and social service agencies (Samet et al., 2004a). Individuals were eligible if they had a documented HIV antibody test by ELISA (confirmed by Western blot) and ≥2 positive answers to the CAGE alcohol screening questionnaire (Mayfield et al., 1974) or an alcohol use disorder diagnosis (lifetime) by a study physician clinical assessment. Participants were fluent in either English or Spanish. Exclusion criteria were cognitive impairment (score of <21 on the Mini-Mental State Examination) and inability to provide informed consent (Folstein et al., 1975; Smith et al., 2006). The study was approved by the Institutional Review Boards of Boston Medical Center and Beth Israel Deaconess Medical Center. Study participants who attended the baseline assessment and at least 2 follow-up visits (total of 3 or more assessments) were included in this analytic sample.

2.1. Outcome

The primary outcome of this study was having a favorable drinking pattern over time. At baseline and at each follow-up visit, alcohol consumption was assessed using a validated calendar method (30-day Timeline FollowBack) (Sobell and Sobell, 1995). Participants were classified at each assessment as abstinent, drinking below or drinking above risky amounts (as defined by the National Institute on Alcohol Abuse and Alcoholism [≥5 drinks/occasion or >14 drinks/week for men; 4 or 7 drinks, respectively, for women and persons aged 65 and over]). Longitudinal drinking patterns were summarized as favorable or unfavorable. A favorable drinking pattern was defined as not drinking risky amounts at every observed study visit (i.e., abstinent or consistently drinking below risky amounts) or a decrease in the observed drinking over time (e.g., from risky to not risky) with no observed increases. All other drinking patterns were defined as unfavorable (i.e. consistently drinking risky amounts, increase in drinking from not risky to risky amounts, or intermittent risky drinking).

In addition, since it is uncertain whether risky amounts as defined in the general population are appropriate definitions for HIV-infected adults, we evaluated the secondary outcome continuous abstinence, defined as reporting abstinence at every study visit.

2.2. Factors associated with drinking patterns

Factors of interest and potential confounders of these associations were assessed at baseline (defined a *priori* based on published literature and clinical experience). Marital status, homelessness (Kertesz et al., 2003), age, gender, race/ethnicity, primary HIV risk factor at the time of infection (injection drug use, men sex with men, heterosexual sex), recent heroin, cocaine and marijuana use (past 12 months), any attendance at Alcoholics Anonymous (AA) meetings (past 6 months), and whether or not the individual spent time with people who drink alcohol (a measure of social pressure to drink) were self-reported. Health-related quality of life was summarized using the Mental Component Summary (MCS) and Physical Component Summary (PCS) scores of the 12-item Short-Form Health Survey (Delate and Coons, 2000). Past 6-month and lifetime diagnosis of alcohol dependence were assessed using the Composite International Diagnostic Interview (World Health Organization, 1996).

We hypothesized that marital status, homelessness, AA attendance, and lack of social pressure to drink would be associated with more favorable drinking pattern; worse health-related quality of life and recent drug use would be associated with an unfavorable pattern. Other variables were considered potential confounders.

2.3. Analysis

We determined the frequency and proportion of each drinking pattern based on the observed data for each person. Multiple logistic regression models were fit to identify baseline factors associated with a favorable alcohol consumption pattern across time. The models adjusted for all factors of interest and potential confounders, none of which were highly correlated (Spearman $r > 0.40$). The Hosmer–Lemeshow test was used to assess model goodness-of-fit. All analyses were conducted using two-sided tests and a significance level of 0.05. Analyses were performed using SAS software (version 9.1; SAS Institute, Cary, NC).

Table 1
Baseline characteristics of a prospective cohort of adults with HIV-infection and current or past alcohol problems [$n = 358$].

Age, mean (SD)	41.6 (7.4)
Female, *n* (%)	89 (24.9)
Race/ethnicity, *n* (%)	
Black	157 (43.9)
White	119 (33.2)
Latino	63 (17.6)
Other	19 (5.3)
Homelessness, *n* (%)	84 (23.5)
Currently married, *n* (%)	22 (6.1)
Mental Component Summary score (MCS), mean (SD)	39.8 (11.4)
Physical Component Summary score (PCS), mean (SD)	43.5 (9.8)
Recent drug use (past 12 months), *n* (%)	
Heroin use	88 (24.6)
Cocaine use	163 (45.5)
Marijuana use	133 (37.2)
Drinks per day, last 30 days, mean (SD)	1.8 (5.5)
Attended AA meeting, past 6 months, *n* (%)	203 (56.7)
Alcohol dependence diagnosis (current), *n* (%)	35 (9.8)
Alcohol abuse diagnosis (lifetime), *n* (%)	61 (17.0)
Alcohol dependence diagnosis (lifetime), *n* (%)	245 (68.4)
Baseline drinking status, *n* (%)	
At risk drinking	114 (31.8)
"Moderate" drinking	41 (11.5)
Abstinent	203 (56.7)
Spending time with people who drink alcohol (Social pressure to drink alcohol), *n* (%)	219 (61.2)
Primary HIV risk factor, at the time of infection, *n* (%)	
Injection drug use	184 (51.5)
Men sex with men	82 (23.0)
Heterosexual sex	91 (25.5)

SD = Standard deviation. Homelessness was defined as at least one night on the street or in a shelter in the past 6 months. AA = Alcoholics Anonymous. Alcohol consumption (drinks per day) was assessed using the Timeline FollowBack method. The presence of a current (past 6 months) alcohol dependence, and a lifetime diagnosis of alcohol abuse or dependence was assessed using the Composite International Diagnostic Interview (CIDI). Risky drinking was defined as >14 drinks/week or ≥5 drinks on an occasion for men, >7 drinks/week or ≥4 drinks on an occasion for women and persons ≥ age 65 years. Moderate drinking was defined as drinking alcohol but below risky drinking limits. A drink was defined as 12–14 g of ethanol, as in the amount in the U.S. in one 12 ounce beer, one 5 ounce glass of wine, or 1.5 ounces of 80 proof liquor. Social pressure to drink alcohol: Subjects were asked how many of the people they spend time with currently drink alcohol (none, a few, about half, most, all). Answers were later dichotomized (none vs. other).

3. Results

Of the 400 participants in the HIV-LIVE prospective cohort study, 358 (90%) completed at least 3 study visits and were included in the present analyses. Median follow-up was 3.4 years. The baseline characteristics of the study sample are presented in Table 1. Over the course of follow-up, 54.5% had a favorable drinking pattern with 43.9% consistently drinking below risky limits at every assessment (70.7% of the latter group abstained), and 10.6% decreasing consumption over time. Of the 45.5% with an unfavorable drinking pattern, 3.6% drank risky amounts at each assessment, 4.7% increased their drinking to risky amounts and 37.2% had intermittent risky drinking.

In the logistic regression model (Table 2), among the factors of interest, recent marijuana use had a negative association with a favorable drinking pattern. Current alcohol dependence (past 6 months) was also negatively associated with an unfavorable pattern. Older age, female gender and having a primary HIV risk factor of injection drug use at the time of infection (compared to heterosexual sex) were associated with a favorable drinking pattern.

Older age and female gender were also associated with continuous abstinence while marijuana use, cocaine use, social pressure to

Table 2
Baseline factors associated with a favorable drinking pattern and continuous abstinence in multivariable logistic regression analyses among a prospective cohort of adults with HIV-infection and current or past alcohol problems [$n = 358$].

	Favorable drinking pattern Adjusted odds ratio (95% CI)	Continuous abstinence Adjusted odds ratio (95% CI)
Baseline factors of interest		
Homelessness	0.72 (0.41, 1.28)	0.61 (0.31, 1.20)
Currently married	0.64 (0.24, 1.70)	0.68 (0.23, 2.03)
Mental Component Summary score (MCS), per 1 point increase	0.99 (0.97, 1.02))	0.98 (0.96, 1.00)
Physical Component Summary score (PCS), per 1 point increase	1.02 (0.99, 1.04)	1.02 (0.99, 1.04)
Recent heroin use (past 12 months)	1.35 (0.70, 2.61)	1.62 (0.76, 3.46)
Recent cocaine use (past 12 months)	0.86 (0.50, 1.47)	**0.39** (**0.20**, **0.74**)
Recent marijuana use (past 12 months)	**0.52** (**0.31**, **0.87**)	**0.53** (**0.29**, **0.99**)
Attending AA meetings, past 6 months	1.37 (0.83, 2.28)	1.40 (0.79, 2.46)
Spending time with people who drink alcohol (social pressure to drink alcohol)	0.81 (0.49, 1.35)	**0.41** (**0.24**, **0.72**)
Confounders		
Age, per 1 year increase	**1.04** (**1.00**, **1.07**)	**1.04** (**1.00**, **1.08**)
Gender (female)	**2.44** (**1.33**, **4.50**)	**3.44** (**1.80**, **6.59**)
Race/ethnicity (reference group: black)		
Hispanic	0.73 (0.37, 1.46)	0.74 (0.35, 1.59)
White	1.15 (0.66, 2.00)	1.12 (0.60, 2.07)
Other	0.68 (0.24, 1.93)	0.50 (0.14, 1.84)
Primary HIV risk factor, at the time of infection (reference group: heterosexual sex)		
Injection drug use	**2.01** (**1.05**, **3.87**)	2.05 (0.96, 4.36)
Men sex with men	1.81 (0.86, 3.80)	1.86 (0.75, 4.61)
Alcohol dependence diagnosis (current)	**0.38** (**0.17**, **0.88**)	**0.32** (**0.11**, **0.99**)

AOR: Adjusted odds ratio from multiple logistic regression analysis. Homelessness was defined as at least one night on the street or in a shelter in the past 6 months. The presence of a current (past 6 months) diagnosis of alcohol dependence was assessed using the Composite International Diagnostic Interview (CIDI). Social pressure to drink alcohol: Subjects were asked how many of the people they spend time with currently drink alcohol (none, a few, about half, most, all). Answers were later dichotomized (none vs. other). The Hosmer and Lemeshow Chi-square test for the multivariable model suggests satisfactory model fit ($p = 0.7$ and $p = 0.9$ for favorable drinking pattern and continuous abstinence, respectively).
Bold values indicate statistically significant associations.

drink and current alcohol dependence were negatively associated with continuous abstinence.

4. Discussion

In a prospective cohort of HIV-infected adults with current or past alcohol problems, almost half of the subjects had an unfavorable drinking pattern over time, but most with an unfavorable pattern were not consistently drinking risky amounts; they varied their drinking between risky and less than risky amounts. Thus an unfavorable drinking pattern is not a fixed state. Most subjects with a favorable drinking pattern were consistently abstinent or consistently drinking less than risky amounts. Unlike the unfavorable drinking pattern, the favorable pattern appears to be far more stable, with consistent abstinence as the most common pattern over time.

Compared with other cohorts of subjects with alcohol use disorders, not specifically with HIV infection, the proportion of abstinent subjects among those with a favorable drinking pattern in this study (>70%) was higher. In general populations, people with alcohol dependence in recovery are evenly spread between abstinence and moderate drinking (Sobell et al., 1996; Dawson et al., 2005). Similar to non-HIV-infected populations, our study found abstinence to be the most stable form of remission (Dawson et al., 2007).

Adjusting for the severity of alcohol problems, we identified various factors independently associated with drinking pattern. Our results add to the evidence linking marijuana use to unfavorable drinking outcomes (both primary and secondary outcomes). Adolescent marijuana use is associated with heavy drinking in adulthood (Merline et al., 2008). Marijuana use is also a risk factor for alcohol dependence (Sartor et al., 2007). Our finding is consistent with studies showing an association between marijuana use and hazardous drinking among people in emergency departments (Woolard et al., 2003). Our findings of associations between age and female gender and favorable drinking patterns are also consistent with the results of general population studies (Dawson et al., 2005). Injection drug use as a risk factor at the time of HIV infection was associated with a lower odd of an unfavorable alcohol consumption outcome. We included this factor as a covariate to control for HIV risk. It was not identified as a predictor of interest as we did not have a hypothesis about its specific impact and thus would caution against over interpretation.

The effects of homelessness, heroin use, cocaine use, social pressure to drink, and AA attendance on the primary outcome were not statistically significant in this cohort although we may have had limited power to detect these associations. Nevertheless, even in the case of limited power, these factors appear to have weaker associations compared to factors of similar prevalence that were associated with drinking patterns. Of note, social pressure to drink and cocaine use was negatively associated with continuous abstinence.

This study has limitations. First, even though the study was prospective and analyses were adjusted for potential confounders, associations may or may not be causal. Second, the cut offs we used for risky drinking were defined for the general population and have not been specifically defined for HIV-infected patients. For a variety of reasons (e.g., susceptibility to hepatic toxicity, medication interactions, immunosuppression) it is likely that among HIV-infected patients, amounts of alcohol that risk health consequences may be lower than in the general population. As such, some participants classified in our study in the favorable drinking pattern group may have been at risk for harm from drinking. However, this misclassification is likely not a major issue in this analysis since most participants with favorable drinking were abstinent. In addition, findings for continuous abstinence were similar with marijuana use, age, gender, cocaine use and social pressure to drink associated with abstinence. Prior literature also suggests an impact of other drug use and environmental factors (e.g., social pressure to drink) on the course of alcohol use disorders (Zywiak et al., 2006a; Moss et al., 2007). Lastly, generalizability of our findings may not extend beyond urban U.S. HIV-infected populations of adults with alcohol problems.

This study also has notable strengths. The most important strength is that we were able to define drinking patterns that reflected changes in alcohol consumption over time, therefore providing a more accurate picture of the drinking patterns in this population compared to cross-sectional studies. Since intermittent risky drinking is common, future research should focus on drinking and consequences over time in HIV-infected adults. To more accurately assess the impact of drinking on clinical outcomes, repeated assessments of drinking over time will be preferable to one-time measures that are unable to capture temporal variability.

Our finding that half of the cohort had a favorable drinking pattern suggests that a favorable evolution of drinking amounts among HIV-infected patients with alcohol problems is common. Clinicians and patients alike should have reason for optimism. Patients should be supported in their efforts to reduce drinking, and success should be acknowledged when it occurs. Although larger studies may identify other factors associated with drinking patterns, those we identified make clinical sense and may be clinically useful for identifying and addressing risky drinking among HIV-infected patients with alcohol problems. Variations over time and the substantial risk of recurrent risky drinking suggest that repeated screening and brief intervention should be done for HIV-infected adults with past alcohol problems, even when patients have not recently exceeded safe drinking limits.

Role of funding source

Support for this study came from the following grants from the National Institute of Alcohol Abuse and Alcoholism (NIAAA) of the National Institutes of Health: 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), and K24-AA015674 (Impact of Alcohol Use on HIV Infection in US and Russia). The NIAAA had no further role in study design; in the collection, analysis and interpretation of data; in the writing of the report; or in the decision to submit the paper for publication. This research was conducted in part at the General Clinical Research Center at Boston University School of Medicine (US Public Health Service [PHS] grant M01 RR00533) and at the Beth Israel Deaconess General Clinical Research Center (USPHS grant M01 RR01032).

Contributors

Nicolas Bertholet, MD: I declare that I have participated in the study design, analyses, interpretation of results, and manuscript writing. Debbie M. Cheng, ScD: I declare that I have participated in the study design, analyses, interpretation of results, and manuscript writing. Jeffrey H. Samet, MD, MA, MPH: I declare that I have participated in the study design, analyses, interpretation of results, and manuscript writing. Emily Quinn, MA: I declare that I have participated in the data management, analyses and manuscript writing. Richard Saitz, MD, MPH: I declare that I have participated in the study design, analyses, interpretation of results, and manuscript writing. All authors contributed to and have approved the final manuscript.

Conflict of interest

All authors declare they have no conflicts of interest.

References

Bertholet, N., Daeppen, J.B., Wietlisbach, V., Fleming, M., Burnand, B., 2005. Reduction of alcohol consumption by brief alcohol intervention in primary care: systematic review and meta-analysis. Arch. Intern. Med. 165, 986–995.

Chander, G., Himelhoch, S., Moore, R.D., 2006. Substance abuse and psychiatric disorders in HIV-positive patients: epidemiology and impact on antiretroviral therapy. Drugs 66, 769–789.

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? J. Acquir. Immune Defic. Syndr. 33, 521–525.

Dawson, D.A., Goldstein, R.B., Grant, B.F., 2007. Rates and correlates of relapse among individuals in remission from DSM-IV alcohol dependence: a 3-year follow-up. Alcohol Clin. Exp. Res. 31, 2036–2045.

Dawson, D.A., Grant, B.F., Stinson, F.S., Chou, P.S., Huang, B., Ruan, W.J., 2005. Recovery from DSM-IV alcohol dependence: United States, 2001–2002. Addiction 100, 281–292.

Delate, T., Coons, S.J., 2000. The discriminative ability of the 12-item short form health survey (SF-12) in a sample of persons infected with HIV. Clin. Ther. 22, 1112–1120.

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. J. Psychiatr. Res. 12, 189–198.

Kaner, E.F., Beyer, F., Dickinson, H.O., Pienaar, E., Campbell, F., Schlesinger, C., Heather, N., Saunders, J., Burnand, B., 2007. Effectiveness of brief alcohol interventions in primary care populations. Cochrane Database Syst. Rev. (2), CD004148.

Kertesz, S.G., Horton, N.J., Friedmann, P.D., Saitz, R., Samet, J.H., 2003. Slowing the revolving door: stabilization programs reduce homeless persons' substance use after detoxification. J. Subst. Abuse Treat. 24, 197–207.

Lefevre, F., O'Leary, B., Moran, M., Mossar, M., Yarnold, P.R., Martin, G.J., Glassroth, J., 1995. Alcohol consumption among HIV-infected patients. J. Gen. Intern. Med. 10, 458–460.

Mayfield, D., McLeod, G., Hall, P., 1974. The CAGE questionnaire: validation of a new alcoholism screening instrument. Am. J. Psychiatry 131, 1121–1123.

Merline, A., Jager, J., Schulenberg, J.E., 2008. Adolescent risk factors for adult alcohol use and abuse: stability and change of predictive value across early and middle adulthood. Addiction 103 (Suppl 1), 84–99.

Moss, H.B., Chen, C.M., Yi, H.Y., 2007. Subtypes of alcohol dependence in a nationally representative sample. Drug Alcohol Depend. 91, 149–158.

Samet, J.H., Cheng, D.M., Libman, H., Nunes, D.P., Alperen, J.K., Saitz, R., 2007a. Alcohol consumption and HIV disease progression. J. Acquir. Immune Defic. Syndr. 46, 194–199.

Samet, J.H., Horton, N.J., Meli, S., Freedberg, K.A., Palepu, A., 2004a. Alcohol consumption and antiretroviral adherence among HIV-infected persons with alcohol problems. Alcohol Clin. Exp. Res. 28, 572–577.

Samet, J.H., Phillips, S.J., Horton, N.J., Traphagen, E.T., Freedberg, K.A., 2004b. Detecting alcohol problems in HIV-infected patients: use of the CAGE questionnaire. AIDS Res. Hum. Retroviruses 20, 151–155.

Samet, J.H., Walley, A.Y., Bridden, C., 2007b. Illicit drugs, alcohol, and addiction in human immunodeficiency virus. Panminerva Med. 49, 67–77.

Sartor, C.E., Lynskey, M.T., Heath, A.C., Jacob, T., True, W., 2007. The role of childhood risk factors in initiation of alcohol use and progression to alcohol dependence. Addiction 102, 216–225.

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.

Sobell, L.C., Cunningham, J.A., Sobell, M.B., 1996. Recovery from alcohol problems with and without treatment: prevalence in two population surveys. Am. J. Public Health 86, 966–972.

Sobell, L.C., Sobell, M.B., 1995. Alcohol Timeline FollowBack (TLFB) Users' Manual. Addiction Research Foundation, Toronto, Canada.

Stein, M.D., Charuvastra, A., Maksad, J., Anderson, B.J., 2002. A randomized trial of a brief alcohol intervention for needle exchangers (BRAINE). Addiction 97, 691–700.

Woolard, R., Nirenberg, T.D., Becker, B., Longabaugh, R., Minugh, P.A., Gogineni, A., Carty, K., Clifford, P.R., 2003. Marijuana use and prior injury among injured problem drinkers. Acad. Emerg. Med. 10, 43–51.

World Health, O., 1996. Composite International Diagnostic Interview (CIDI) (Core Version 2.0). World Health Organization, Geneva, Switzerland.

Zywiak, W.H., Stout, R.L., Longabaugh, R., Dyck, I., Connors, G.J., Maisto, S.A., 2006a. Relapse-onset factors in Project MATCH: the relapse questionnaire. J. Subst. Abuse Treat. 31, 341–345.

Zywiak, W.H., Stout, R.L., Trefry, W.B., Glasser, I., Connors, G.J., Maisto, S.A., Westerberg, V.S., 2006b. Alcohol relapse repetition, gender, and predictive validity. J. Subst. Abuse Treat. 30, 349–353.

Interacting with Computers 22 (2010) 289–298

Contents lists available at ScienceDirect

Interacting with Computers

journal homepage: www.elsevier.com/locate/intcom

Response to a relational agent by hospital patients with depressive symptoms

Timothy W. Bickmore [a,*], Suzanne E. Mitchell [b], Brian W. Jack [b], Michael K. Paasche-Orlow [b], Laura M. Pfeifer [a], Julie O'Donnell [b]

[a] College of Computer and Information Science, Northeastern University, 360 Huntington Ave, WVH202, Boston, MA 02115, USA
[b] Boston University School of Medicine/Boston Medical Center, 850 Harrison Ave, 2nd Floor, Boston, MA 02118, USA

ARTICLE INFO

Article history:
Received 19 July 2009
Received in revised form 23 October 2009
Accepted 6 December 2009
Available online 8 January 2010

Keywords:
Depression
Relational agent
Embodied conversational agent
Hospital discharge
Therapeutic alliance

ABSTRACT

Depression affects approximately 15% of the US population, and is recognized as an important risk factor for poor outcomes among patients with various illnesses. Automated health education and behavior change programs have the potential to help address many of the shortcomings in health care. However, the role of these systems in the care of patients with depression has been insufficiently examined. In the current study, we sought to evaluate how hospitalized medical patients would respond to a computer animated conversational agent that has been developed to provide information in an empathic fashion about a patient's hospital discharge plan. In particular, we sought to examine how patients who have a high level of depressive symptoms respond to this system. Therapeutic alliance – the trust and belief that a patient and provider have in working together to achieve a desired therapeutic outcome – was used as the primary outcome measure, since it has been shown to be important in predicting outcomes across a wide range of health problems, including depression. In an evaluation of 139 hospital patients who interacted with the agent at the time of discharge, all patients, regardless of depressive symptoms, rated the agent very high on measures of satisfaction and ease of use, and most preferred receiving their discharge information from the agent compared to their doctors or nurses in the hospital. In addition, we found that patients with symptoms indicative of major depression rated the agent significantly higher on therapeutic alliance compared to patients who did not have major depressive symptoms. We conclude that empathic agents represent a promising technology for patient assessment, education and counseling for those most in need of comfort and caring in the inpatient setting.

1. Introduction

Depression is a common, debilitating mental health disorder and is one of the leading causes of disability among men and women of all ages worldwide (World Health Organization, 2001). Although depression is treatable, stigma associated with mental health problems in general, and depression in particular, represents a significant barrier to treatment (Sirey et al., 2001). In addition, access to mental health care can also present a formidable barrier to depression treatment, particularly in rural areas of the country (US Department of Health and Human Services, 2004). Furthermore, once patients are in treatment for depression, relapses due to non-adherence are common.

Computerized systems hold the promise of increasing the reach and efficacy of depression screening and treatment interventions. Several of these systems have now been developed and successfully evaluated, demonstrating efficacy for the identification and treatment of depression (Fann et al., 2009; Marks et al., 2003; Wright et al., 2005). Central to the effective treatment of depression is the therapeutic alliance – the collaborative bond between patient and mental health provider (Krupnick et al., 2006), and we would expect that the therapeutic alliance is equally important to the efficacy of computerized depression treatment programs (Bickmore et al., 2005b). Yet, existing automated systems for detecting and treating depression have not focused on the development of therapeutic alliance.

Over the last 4 years we have been developing and evaluating a computer animated conversational agent, designed to provide information and counseling to hospital patients at the time of hospital discharge. The agent has been designed with many verbal and nonverbal behaviors intended to foster the development of a therapeutic alliance with patients, such as empathy and social dialogue. In the current study, we sought to explore how patients with major depressive symptoms would react to the agent. Specifically, we investigated whether they would find it more or less acceptable and usable, and whether they would form a stronger or weaker therapeutic alliance with the agent, compared to patients without major depressive symptoms. Positive findings would indicate that animated conversational agents with relational behavior could provide an effective medium for delivering automated depression screening and treatment.

* Corresponding author. Tel.: +1 617 373 5477; fax: +1 617 812 2589.
E-mail address: bickmore@ccs.neu.edu (T.W. Bickmore).

0953-5438/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.intcom.2009.12.001

1.1. Depression

Depression affects 13–16% of people in the US, and is an important risk factor for functional impairment and poor outcomes among patients with various chronic illnesses such as diabetes, coronary artery disease and multiple sclerosis (Hasin et al., 2005; Parashar et al., 2006; Scherer and Herrmann-Lingen, 2009). Depression is associated with both physiological and behavioral factors affecting health outcomes for chronic disease patients including poor self-care management, reduced treatment plan adherence, and certain medical conditions such as chronic inflammatory states and hypercoagulability.

Despite the availability of depression treatment, less than one third of depressed patients receive treatment (Pratt and Brody, 2008). While depressed patients often express interest in treatment they also report barriers including time constraints, stigma, childcare conflicts, lack of transportation, and poor access to mental health services particularly in rural areas (Goodman, 2009). Studies also show that when physicians face a burden of competing clinical priorities for a given medical encounter, depression often goes untreated. Other barriers cited by physicians include fragmented mental health systems, lack of insurance coverage, patient resistance and difficulty making the diagnosis of depression.

1.2. Health literacy and mental health

Health literacy is also a central interest of ours in developing agents for inpatient education and counseling, given the low levels of health literacy in our patient population (Paasche-Orlow et al., 2005). Health literacy is the ability to perform the basic reading and numerical tasks required to function in the healthcare environment (American Medical Association Ad Hoc Committee on Health Literacy for the Council on Scientific Affairs, 1999). Health literacy is not simply the ability to read; it also requires a complex set of analytical and decision-making skills, and the ability to apply these skills to health situations. Fully 36% of American adults have limited health literacy skills, with even greater prevalence among patients with chronic diseases, those who are older, and those who have lower levels of education (Paasche-Orlow et al., 2005). Among indigent and minority patients in urban areas this number rises to over 80% (Williams et al., 1995).

A handful of studies have evaluated the relationship between health literacy and mental illness (Francis et al., 2007; Gazmararian et al., 2000; Lincoln et al., 2006, 2008; Morris et al., 2006; Weiss et al., 2006). In most of these studies limited health literacy has been associated with higher rates of depressive symptoms. For example, one report demonstrated that health literacy has an important longitudinal relationship with the course of depressive symptoms among adults with addiction (Lincoln et al., 2006). Given this high degree of association between depression and health literacy, interventions targeted at assessing and treating depression must be designed to accommodate patients with limited health literacy.

1.3. Relational agents in mental health

Relational agents – animated conversational agents designed to establish trust and therapeutic alliance with users over time – represent a potentially powerful technology for delivering health care services to patients with mental illness. Fig. 1 shows the relational agent interface used in our work. These interfaces use the easy-to-understand format of face-to-face conversation, making them less intimidating and more accessible to patients with a wide range of computer, reading and health literacy skills, and a low-pressure environment where patients are free to take as much time as they need. In prior work, this interface has been found to be easily used by a wide range of participants, including those with no prior experience with computers (Bickmore et al., 2005a). In addition, these agents use many verbal and nonverbal behaviors designed to establish a therapeutic alliance with patients, including displays of empathy, close proximity, more frequent eye gaze and attentiveness, and social dialogue and humor (Bickmore et al., 2005b). Finally, many studies have shown that patients are more honest with a computer than a human clinician when disclosing potentially stigmatizing behaviors such as alcohol consumption and HIV risk behavior (Ahmad et al., 2009; Card and Lucas, 1981; Ghanem et al., 2005; Kissinger et al., 1999; Newman et al., 2002). Individuals with depression may thus find a relational agent more approachable than a clinician in many situations, making it more effective at depression screening and counseling.

1.4. Overview

The development of the relational agent followed work over the last decade on a new standardized hospital discharge protocol de-

Fig. 1. Relational agent interface for virtual hospital discharge nurse application.

signed to reduce re-hospitalizations. A key element of this protocol involves a human nurse spending half an hour with each patient, ensuring they understand their post-discharge self-care regimen before they leave the hospital. While this protocol was shown to be effective in a previous clinical trial (Jack et al., 2009), the additional nursing time required made it impractical to disseminate, motivating the development of the relational agent to perform this task. A more detailed description of the agent development methodology and pilot test results was presented in Bickmore et al. (2009).

In the rest of this paper we first review related work, then briefly review the design of our relational agent for hospital patient education. We then describe a new study assessing the acceptance of the agent by hospitalized medical patients with and without symptoms of depression, as part of a new, ongoing clinical trial.

2. Related work

2.1. Technology in mental health

Internet-based technology has been explored extensively for opportunities to expand access to mental health education and services. Mental health internet sites are now used to provide information, screen for mental health and mood disorders, assist in the delivery of treatment, and provide social and therapeutic support. Web-based interventions are effective for a range of mental health disorders including depression, panic, post-traumatic stress disorder, perceived stress in schizophrenia, stress, insomnia, and eating disorders. Overall, computer-based mental health services are well received by patients and clinically useful in feasibility trials, although an actively depressed mood can be a barrier to use.

2.2. Technologies for depression intervention and treatment

Investigators have reported mixed results from clinical trials of internet-based depression screening and treatment programs (Christensen et al., 2004; Clarke et al., 2002, 2005). In general, computer-based technology has been used to substitute or augment traditional face-to-face therapeutic contact and the delivery of self-help materials. Studies using internet technology for depression support groups have shown reduced social isolation among users and perceived benefit from participation (Hill et al., 2006; Weinert et al., 2008). However, some reports indicate that adolescents, who participated in internet support groups, had reduced interpersonal communication with family members. Screening for depression using internet-based self-assessment tools has proved successful, although minority and elderly persons are less likely to engage in web-based depression screening programs (Houston et al., 2001). Internet-based Cognitive Behavior Therapy and some self-help depression interventions are effective in relieving depression symptoms in mild to moderately depressed patients, and in some cases even more effective with severely depressed patients (Houston et al., 2001). Among factors positively correlated with computer use for mental health treatment are perceived usefulness of the treatment, preference for anonymity, ease of contact and ease of use (Lai et al., 2008). However, according to one study, when intervention programs are accessed outside the context of a randomized research trial, attrition rates are high. This suggests that contact with a mental health provider may be necessary.

While currently available internet-based depression treatment can improve access to care, new technologies that serve as third parties to therapeutic interventions are on the horizon to enhance patient engagement in mental health care. For example, Coyle and Doherty described a 3D computer game designed to enhance adolescent engagement in therapeutic discussion (Doherty and Coyle, 2009). In their study, therapists reported overall acceptance of the game therapy by patients, however no evidence on patient preference or attitudes toward the technology was available from the subjects themselves.

Lisetti used animated anthropomorphic conversational agents to address social phobias such as public speaking anxiety, panic disorder and agoraphobia, in which the patient interacts with the agent in the reenactment of a fear-inducing situation (Lisetti, 2008).

Patients participating in internet or computer-based depression interventions also face some challenges. Some of the barriers reported by users include time constraints, lack of motivation, technical or computer-access problems, physical illness, the lack of face-to-face contact, preference for taking medication, perceived lack of treatment effectiveness, and burden of the program (Christensen et al., 2009). One commonly encountered barrier to computer interaction for depression across trials is active depressive symptoms (Doherty and Coyle, 2009). For example, in the *T*ailored *I*nterventions for Management of *DE*pressive *S*ymptoms (TIDES) program, a computer-based education program on self-care strategies for depressive symptoms in persons living with HIV/AIDS, was rated as easy to use and useful, but computer anxiety and depressive symptoms were negatively correlated with intention to use. Klein et al. reported similar difficulties in a trial of cognitive behavior therapy for panic disordered patients (Klein et al., 2006).

2.3. Conversational agents in mental health

Some of the earliest dialogue systems developed in healthcare were designed for psychotherapy applications. The ELIZA system was developed in 1966 to simulate the behavior of a Rogerian psychotherapist, in which the patient and the computer exchanged typed text messages (Weizenbaum, 1966). Although ELIZA was not intended to be used for actual therapy, similar systems have been proven effective for therapy in which the system is essentially prompting a patient to think aloud and work through his or her own problems (Slack, 2000). In these applications, significant errors in understanding user input or in producing incoherent system output can often be tolerated, as the primary function of the system is just to keep the user engaged in the interaction.

Colby developed an ELIZA-like system that was designed to use Cognitive Behavioral Therapy to treat individuals with depression. In addition to providing typed text counseling with patients, the system provided text-based educational materials about depression (Colby, 1995). While Colby reported that the program was accepted by patients, evaluations by other researchers indicate the typed text medium confused some patients and the only comparative evaluation in the literature indicates that the system did not work as well as clinician-administered therapy (Wright, 2004).

2.4. Relational agents in mental health

Bickmore developed a relational agent to promote medication adherence among patients with schizophrenia. The agent tracks each patient's medication taking behavior for a single antipsychotic medication taken by mouth in pill or capsule form based on self-report, and also promotes physical activity (walking), and talking to the agent every day. For each of these three behaviors, the agent first asks for a self report of behavior, provides feedback on the behavior, and negotiates a behavioral goal. Intervention on each behavior is started and terminated according to a schedule for a 30-day intervention. Several elements were incorporated into the system to address the needs of individuals with schizophrenia, including extended orientation and termination of the therapeutic relationship, use of concrete language, and certain nonverbal behaviors. A 30-day quasi-experimental pilot study involving 20 patients indicated high levels of acceptance, usability, and self-reported adherence (Bickmore and Pfeifer, 2008).

3. Design of a relational agent for patient education at hospital discharge

We have developed an automated system that teaches hospital patients about their post-discharge self-care regimen, including medications, follow-up appointments, exercise and diet regimens, and pending lab tests. The system is designed to be used by patients while they are still in their hospital beds. In order to make the system as acceptable and effective as possible, we designed the interface to incorporate an animated virtual nurse agent who embodies best practices in health communication for patients with inadequate health literacy. The agent is deployed on a wheeled kiosk with a touch screen display attached to an articulated arm that can be positioned in front of patients while they are in bed (Fig. 1). The agent is designed to interact with patients once every day they are in the hospital, but the primary interaction is just prior to hospital discharge. At this time, the patient spends approximately half an hour using the system, reviewing the layout and contents of a personalized "After Hospital Care Plan" (AHCP) booklet that is produced for them and contains their post-discharge self-care instructions. The paper booklet is given to patients before their conversation with the agent, and the agent reviews a digital version of the booklet in the interface, so that patients can follow along with the agent's explanation in their paper booklets.

3.1. Development methodology

Our multi-disciplinary design team comprised HCI researchers, doctors and nurses, a health literacy expert, and programmers and animators. The design process, from project start to completion of the user studies, lasted 3 years.

We used a multi-faceted approach to designing a system that would effectively teach patients, including those with inadequate health literacy, about their hospital discharge instructions. We began our design process with an ethnographic study of a re-engineered hospital discharge intervention that was currently underway in the hospital. Members of the design team visited hospital rooms, attended rounds with the medical team, observed discharge sessions in which nurses taught patients about their AHCP booklets, and interviewed the nurses who were performing this task. From these activities we learned about the stakeholders and basic workflow requirements of the system.

In addition to these "big picture" activities, we investigated the micro-behavior of expert nurses during discharge consultations with patients. We videotaped several mock discharge interactions in which one of the nurses explained an AHCP booklet either to a member of the research staff or a participant recruited from the community. We conducted discourse analyses of the videotaped interactions to characterize the verbal and nonverbal behavior used by the nurses while explaining a booklet to a patient. We also conducted two rounds of user testing of an agent explaining discharge instructions to users in our HCI lab (Bickmore et al., 2008).

3.2. Implementation

The agent was developed using an existing framework for ECA-based health counseling (Bickmore and Picard, 2005), extended with a computational model for the explanation of documents (Bickmore et al., 2008). In this interface, the agent speaks, using a synthetic voice, and displays animated nonverbal behavior (hand gestures, posture shifts, facial expressions, etc.) in synchrony with the speech. User contributions to the conversation are made by touching utterance option buttons on a touch screen display that are dynamically updated for each user speaking turn (Fig. 1). We considered using speech recognition as the input modality rather than the touch screen, but the hospital room can be a very noisy environment, and a significant portion of the patient population speaks English as a second language with many accents that would be problematic for commercial speech recognition products.

Dialogues are scripted, using a custom hierarchical transition network-based scripting language, and a visual dialogue design tool. The final system contains 550 dialogue states including 322 unique medication scripts covering 2254 medicines, along with 48 scripts for diagnoses.

The importance of caring, empathy and good "bedside manner" is widely recognized in healthcare as a key factor in improving not only patient satisfaction, but treatment outcomes across a wide range of health care disciplines (Garrity, 1981), and particularly in nursing (Sourial, 1997). Given this, prior successful implementations of empathic computer agents (Bickmore and Picard, 2005), and the need for the agent to maintain patients' attention for the hour it may take to relate all of the information in their discharge plans, we augmented the agent's informational dialogue with relational dialogue. Following earlier work on relational agents, we integrated a range of relational behavior into the agent dialogue, including appropriate forms of address (calling the patient by name), social chat at the beginning of every interaction, meta-relational communication, appropriate humor, appropriate feedback at every empathic opportunity, and references to information discussed in past interactions to give a sense of continuity (Bickmore et al., 2005a). The agent also offers patients the opportunity to take breaks at several points during the interaction in order to sustain attention and engagement.

A fragment of a typical conversation is shown in Fig. 2. The agent proceeds through the AHCP booklet linearly, describing each section before moving onto the next. Conversations generally consist of: (1) a greeting and social chat (Fig. 3); (2) orientation to the virtual nurse and the discharge process; (3) introduction of the AHCP (Fig. 4); (4) review of medications, including comprehension tests (Fig. 5); review of appointments, including comprehension tests; review of recommended diet and physical activity; "patient activation", in which the patient is urged to keep track of any ques-

VN: Good afternoon. My name is Elizabeth and I am here to help you with the discharge process here at Boston Medical Center. You are Joe, is that right?

Patient: OK

VN: It is great to meet you Joe. As you might have noticed, hospitals are very busy places. My job is to make sure that before you leave, you feel that you have been given all the information you need to help make the transition from here to your home as smooth as possible.

VN: So, are you a Red Sox fan?

… several turns of social chat …

VN: It looks like you are going to be leaving the hospital soon. I know you must be looking forward to leaving the hospital, but before you go, I want to review some information with you.

Patient: OK

VN: *[displays AHCP cover page, gestures with whole hand]* This is a care plan that we created just for you. Remember to bring this booklet with you to all of your doctors appointments.

…

Fig. 2. Sample dialogue – relational aspects highlighted.

Fig. 3. Greeting and social chat with virtual nurse.

tions or issues they want to discuss with their primary care provider (Fig. 6); and their primary diagnosis (Fig. 7). Following the interaction with the system, typically lasting 30–40 min, a report is printed for the human nurse that describes issues raised by the program that require follow up, such as questions about medications that the virtual nurse could not answer.

3.3. Pilot studies

Two rounds of pilot studies were conducted to assess acceptance, usability and satisfaction with the system (Bickmore et al., 2009). Results indicate that patients found the system easy to use, reported high levels of satisfaction, and most said they preferred receiving the discharge information from the agent over their doctor or nurse. Patients also expressed appreciation for the time and attention provided by the virtual nurse, and felt that it provided an additional authoritative source for their medical information.

4. Acceptance of relational agent by patients with depressive symptoms

We sought to explore how hospital patients with major depressive symptoms would react to the relational agent, by investigating whether they would find it more or less acceptable and usable, and whether they would form a stronger or weaker therapeutic alliance with the agent, compared to patients without major depressive symptoms.

Fig. 4. Introduction of after hospital care plan by virtual nurse.

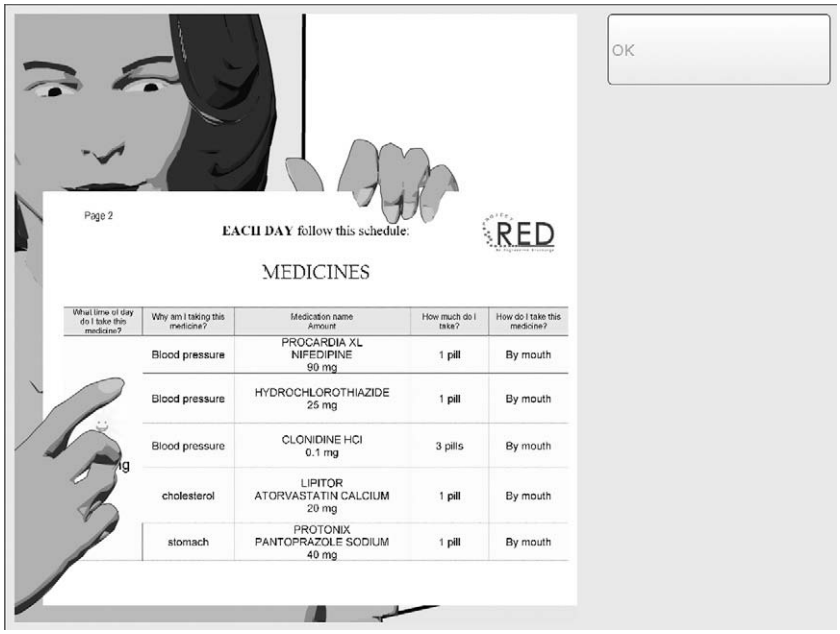


Fig. 5. Review of medications by virtual nurse.

4.1. Methods

A secondary analysis was performed using data from 139 English-speaking, hospitalized adults from an ongoing randomized controlled trial conducted at an urban academic safety-net hospital. The parent study, currently ongoing, is a two-armed evaluation study of the impact of the agent and other improvements to the hospital discharge process on 30-day hospital readmissions.

4.1.1. Study setting

The deployment site for the relational agent system is Boston Medical Center, a 547 bed safety-net hospital that serves an urban, 84% minority, traditionally underserved population. Approximately 58% of this population has inadequate health literacy (see Section 4.1.3 for the measure used).

4.1.2. Participants

Participants in the study were English-speaking adult patients, 18 years or older, admitted to the teaching service of Boston Medical Center, a large urban safety-net hospital with an ethnically diverse patient population. Three hundred and forty-seven subjects were enrolled and randomized between October 15, 2008 and June 20, 2009. Patients were required to have a telephone, be able to comprehend study details and the consent process in English, and have plans to be discharged to a US community. Patients were not enrolled if they were admitted from a skilled nursing facility or

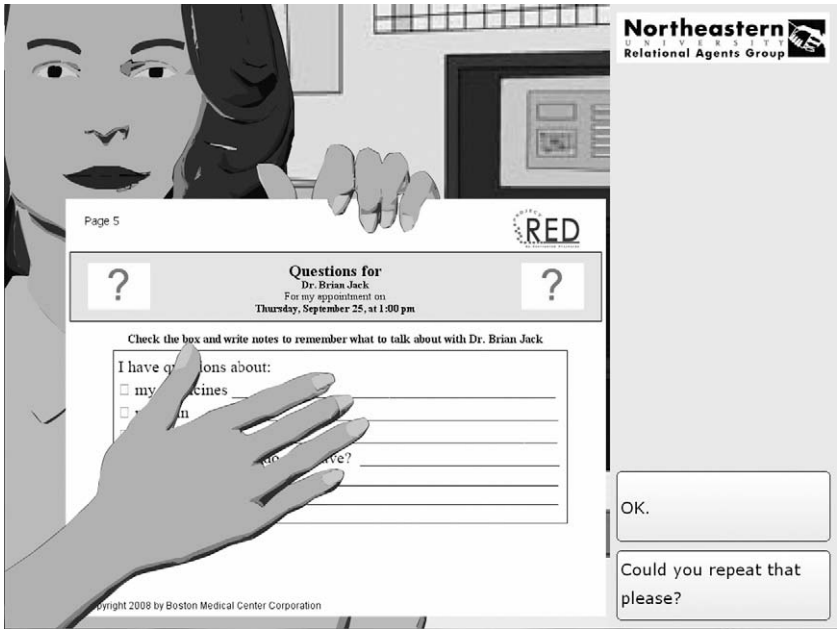


Fig. 6. Patient activation by virtual nurse.

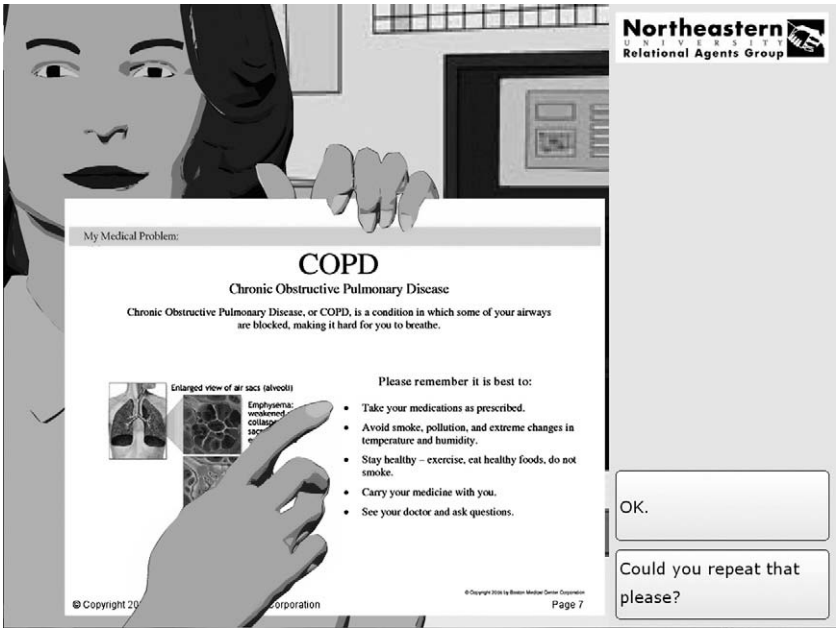


Fig. 7. Review of primary diagnosis by virtual nurse.

other hospital, transferred to a different hospital service prior to enrollment, admitted for a planned hospitalization, on hospital precautions, on suicide watch, deaf or blind.

Of the 347 subjects enrolled into the parent study, 173 were randomized into the relational agent intervention arm of the study. Of these, 131 completed all measures necessary for our analyses.

4.1.3. Measures

4.1.3.1. Depressive symptoms. The primary independent variable of interest was depressive symptoms, defined as a positive score for major depression on the validated PHQ-9 depression screening tool. A dichotomized variable was created using a standardized scoring system to determine the screening cut-off for major depressive symptoms (score $\geqslant$10 on a possible score range of 0–27) (Kroenke et al., 2001).

4.1.3.2. Therapeutic alliance. Patient perception of therapeutic alliance with the agent was assessed using the Bond subscale of the Working Alliance Inventory, a 12-item questionnaire assessing the emotional dimension of a patient's trust and belief that they can work together with their provider to achieve desired therapeutic outcomes (Horvath and Greenberg, 1989).

4.1.3.3. Health literacy. We assessed health literacy using the 66-word version of the Rapid Estimate of Adult Literacy in Medicine (REALM) (Davis et al., 1993). We defined limited health literacy as a reading level of 8th grade and below and adequate health literacy as 9th grade and above for our analyses, as prior authors have done (Lincoln et al., 2006; Lindau et al., 2006; Mancuso and Rincon, 2006; Sudore et al., 2006).

4.1.3.4. Attitudes towards the agent. In addition to therapeutic alliance, we assessed additional patient attitude towards and satisfaction with the agent using single item, scale response questions, shown in Table 1.

4.1.3.5. Questions asked. All patient interactions with the agent were logged for subsequent analysis. From these logs we counted the number of times each patient selected a response that provided additional information when offered as an option.

4.1.4. Procedure

Following enrollment, collection of background, demographic and depressive symptom information and randomization, intervention patients have their post-discharge self-care information entered into a workstation by a study nurse. This information is used to generate the AHCP booklet, and is also downloaded to a mobile kiosk that is then wheeled into the patient's room (Fig. 1). After the patient is given their paper booklet and provided with a brief training session on how to use the touch–screen interface, they are left to conduct their conversation with the agent. At the end of this interaction, any unresolved patient questions or issues are displayed for a human nurse to follow up with the patient, and results of the session are uploaded to a database. At this time, self-report measures covering attitudes towards the agent and therapeutic alliance are collected. All self-report measures are verbally collected to accommodate patients with limited literacy.

Table 1
Self-Report Measures of Attitudes Towards the Relational Agent.

Measure	Question	Anchor 1	Anchor 7
Satisfaction	How satisfied were you with Elizabeth?	Not at all	Very satisfied
Usability	How easy was talking to Elizabeth?	Easy	Difficult
Continue	How much would you like to continue working with Elizabeth?	Not at all	Very much
Relationship	How would you characterize your relationship with Elizabeth?	Complete stranger	Close friend
Preference	Would you rather have talked to your doctor or nurse than Elizabeth?	Definitely prefer doctor or nurse	Definitely prefer Elizabeth
Adherence	How likely is it that you will follow Elizabeth's advice?	Not at all likely	Very likely

Table 2
Patient demographics by depression status.

Characteristics	No major depressive condition, $n = 112$	Major depressive condition, $n = 19$	p-Value
Gender, n (%)			
Male	62 (56)	8 (42)	0.32
Female	49 (44)	11 (58)	
Age, mean (SD)	49 (13)	46 (13)	0.28
Education level, n (%)			
Less than high school	4 (4)	2 (11)	0.41
Some high school	23 (21)	6 (32)	
HS graduate or GED	39 (35)	7 (37)	
Some college	24 (21)	2 (11)	
4-year college graduate or above	19 (17)	2 (11)	
Health literacy level, n (%)			0.09
Grade 3 and below	13 (11)	5 (26)	
Grade 4–6	10 (9)	4 (21)	
Grade 7–8	31 (28)	3 (16)	
Grade 9 and above	58 (52)	7 (37)	
English primary language at home, n (%)	99 (90)	17 (89)	0.94
Patient married, n (%)	21 (19)	1 (5)	0.14
Computer literacy, mean (SD)	2.45 (0.87)	2.0 (1.05)	<0.001
Attitude towards computers, mean (SD)	2.59 (0.81)	2.63 (0.96)	0.13
Have you ever been told by a doctor or therapist that you have depression?	34 (30)	15 (79)	<0.001
Have you ever been prescribed medication for depression? (only if "yes" to above – diagnosed with depression at some point)	29 (26)	14 (74)	<0.001
Do you take medication for depression now? (only if "yes" to above = diagnosed with depression at some point)	12 (11)	9 (47)	<0.001
Length of hospital stay in days, mean (SD)	2.16 (2.99)	2.84 (1.46)	0.13

Table 3
Self-report assessments of the relational agent p values from Mann–Whitney test ($p < 0.05$ in bold).

Measure	No major depressive condition		Major depressive condition		p-Value
	Mean	SD	Mean	SD	
Satisfaction	6.50	1.07	6.59	1.18	0.37
Usability	1.99	1.82	1.71	1.83	0.49
Continue	5.58	1.78	6.56	1.03	**0.03**
Relationship	4.80	1.73	5.24	1.89	0.30
Preference	4.16	2.03	4.73	1.67	0.29
Adherence	6.22	1.21	6.73	0.59	0.13

4.2. Results

All variables were tested for normality. Therapeutic alliance was found to have a significant negative skew and was logarithmically transformed (after reflection) to normalize the distribution. Questions Asked had a positive skew but was not transformed due a floor effect and the nature of the measure (non-parametric tests were used).

Of the 131 patients analyzed, 19 (14.5%) were classified as having major depressive symptoms according to the PHQ-9. Table 2 compares patient demographic and other characteristics by depression status. Patients with major depressive symptoms were similar to other patients on all characteristics except for computer literacy, with the depressive group scoring significantly lower.

Table 3 shows self-report ratings of the agent. Overall ratings of satisfaction and ease of use were very high, and only 24% of patients said they would have preferred receiving their discharge information from their doctor or nurse (40% were neutral, 36% said they definitely preferred the agent). Patients with major depressive symptoms stated a significantly greater desire to continue interacting with the agent, $p < 0.05$ by Mann–Whitney test (mean of 6.6 vs. 5.6).

Table 4 shows correlations among continuous measures. There are significant positive correlations between therapeutic alliance,

Table 4
Correlations among continuous measures.

Measure	Alliance	Literacy	Questions	Satisfy	Usability	Continue	Relationship	Preference	Adherence
Alliance	1	−0.123	0.208*	0.317**	−0.249**	0.408**	0.393**	0.394**	0.319**
Literacy	−0.123	1	−0.193*	−0.093	−0.072	−0.073	−0.069	−0.090	−0.053
Questions	0.208*	−0.193*	1	0.039	0.103	0.208*	0.217*	0.128	0.154
Satisfy	0.317**	−0.093	0.039	1	−0.306**	0.430**	0.253**	0.198*	0.118
Usability	−0.249**	−0.072	0.103	−0.306**	1	−0.166	−0.087	−0.023	−0.223*
Continue	0.408**	−0.073	0.208*	0.430**	−0.166	1	0.497**	0.310**	0.297**
Relationship	0.393**	−0.069	0.217*	0.253**	−0.087	0.497**	1	0.280**	0.171
Preference	0.394**	−0.090	0.128	0.198*	−0.023	0.310**	0.280**	1	0.198*
Adherence	0.319**	−0.053	0.154	0.118	−0.223*	0.297**	0.171	0.198*	1

* $p < 0.05$ level.
** $p < 0.01$.

number of questions asks, satisfaction with the overall virtual nurse system, desire to continue using the system, preference for the virtual nurse over a human doctor or nurse, and stated expectation of following the agent's advice.

Patients with major depressive symptoms scored the agent significantly higher on therapeutic alliance compared to patients classified as not having major depressive symptoms (6.2 vs. 5.5, before transformation), $t(108) = 2.02$, $p < 0.05$, $d = 0.58$.

Patients classified as having inadequate health literacy scored significantly higher on therapeutic alliance compared to patients with adequate health literacy (5.9 vs. 5.4, before transformation), $t(116) = 2.56$, $p < 0.05$, $d = 0.47$. Patients with inadequate health literacy also asked the agent significantly more questions, compared to patients with adequate health literacy, $p < 0.05$ by Mann–Whitney test.

4.3. Discussion

Self-report ratings of satisfaction, ease of use, and attitudes towards the agent were high for all patients, with only 24% of patients indicating they would have preferred receiving their discharge information from their doctor or nurse. This result is similar to that found in our pilot studies, in which patients stated that they appreciated the amount of information given to them by the agent, the amount of time that the agent spent with them, and that the agent did not talk down to them as many providers do (Bickmore et al., 2009). As one patient reported:

> "It was just like a nurse, actually better, because sometimes a nurse just gives you the paper and says 'Here you go.' Elizabeth explains everything."

The primary finding of the study is that patients with symptoms indicative of major depression rated the agent significantly higher on therapeutic alliance compared to patients who did not have major depressive symptoms. In combination with their greater stated desire to continue working with the agent, this indicates that a relational agent is not only acceptable to patients with major depressive symptoms, but that these patients feel they have established a stronger emotional bond with the agent compared to patients without depressive symptoms.

Patients with inadequate health literacy also reported a significantly greater therapeutic alliance with the agent compared to patients with adequate health literacy. This effect was independent of depression. Given that both health literacy and depression represent barriers to healthcare, this indicates that automated patient education systems incorporating relational agent technology could help reduce disparities in access to care.

4.3.1. Limitations

There are several important limitations to our preliminary study. First, our results may not be generalizable to populations other than those served by urban safety-net hospitals or other populations excluded from the parent study. Second, our results are correlational, so we do not know the direction of the associations between major depressive symptoms and other measures reported. Finally, our results may be due to the fact that patients with depressive symptoms react more positively to computer-based interventions, or any novel intervention. This would be consistent with reports from evaluations of many other computer interventions for patients with mental health problems (Wright, 2004). Our observation that patients with inadequate health literacy also had higher therapeutic alliance may serve to partially mitigate this concern.

5. Conclusion and future work

Depressive symptoms are a risk factor for many adverse health conditions and early hospital readmission, thus the ability to identify hospitalized patients for depression is of great interest to the healthcare community. A relational agent that can screen and develop a therapeutic alliance may also enhance the likelihood of successful outpatient depression treatment following hospital discharge. Further, the deleterious effects of active depression on treatment response and disease severity may be ameliorated with prompt identification and initiation of treatment at the time of hospitalization for patients with depressive symptoms.

Low health literacy contributes to health disparities across diverse realms of healthcare, including access to care, medical decision-making, medication adherence, and preparation for a diagnostic test. The significantly greater therapeutic alliance reported by patients of low health literacy may be related to the agent's ability to deliver information in a nonjudgmental manner at a patient-centered pace of learning, alleviating the stigma felt by patients with low health literacy. Developing similar patient counseling applications in other healthcare settings where health literacy impacts disease outcomes could have a positive impact on health disparities.

Our future work in this area involves developing and evaluating relational agents for depression screening and treatment, to determine if the results found in this study do in fact imply the promise of the technology for reaching and treating depressed individuals in the inpatient setting.

Acknowledgements

This work was supported by grants from the NIH National Heart, Lung and Blood Institute and the Agency for Health care Research and Quality. Thanks to the many members of the Re-Engineered Discharge project team at Boston Medical Center and the Relational Agents Group at Northeastern University for their contributions to this work.

References

Ahmad, F., Hogg-Johnson, S., Stewart, D., Skinner, H., Glazier, R., Levinson, W., 2009. Computer-assisted screening for intimate partner violence and control: a randomized trial. Ann. Intern. Med. 151 (2), 93–102.

American Medical Association Ad Hoc Committee on Health Literacy for the Council on Scientific Affairs, 1999. Health literacy: report of the council on scientific affairs. JAMA 281 (6), 552–557.

Bickmore, T., Pfeifer, L., 2008. Relational agents for antipsychotic medication adherence. In: Proceedings of the CHI'08 Workshop on Technology in Mental Health, Florence, Italy.

Bickmore, T., Picard, R., 2005. Establishing and maintaining long-term human-computer relationships. ACM Trans. Comput. Hum. Interact. 12 (2), 293–327.

Bickmore, T., Caruso, L., Clough-Gorr, K., Heeren, T., 2005a. "It's just like you talk to a friend" – relational agents for older adults. Interact. Comput. 17 (6), 711–735.

Bickmore, T., Gruber, A., Picard, R., 2005b. Establishing the computer-patient working alliance in automated health behavior change interventions. Patient Educ. Counsel. 59 (1), 21–30.

Bickmore, T., Pfeifer, L., Yin, L., 2008. The role of gesture in document explanation by embodied conversational agents. Int. J. Semantic Comput. 2 (1), 47–70.

Bickmore, T., Pfeifer, L., Jack, B.W., 2009. Taking the time to care: empowering low health literacy hospital patients with virtual nurse agents. In: Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI), Boston, MA.

Card, W., Lucas, R., 1981. Computer interrogation in medical practices. Int. J. Man Mach. Stud. 14, 49–57.

Christensen, H., Griffiths, K.M., Jorm, A.F., 2004. Delivering interventions for depression by using the internet: randomised controlled trial. Br. Med. J., 328.

Christensen, H., Griffiths, K., Farrer, L., 2009. Adherence in internet interventions for anxiety and depression. J. Med. Internet Res. 11 (2), e13.

Clarke, G., Reid, E., Eubanks, D., O'Connor, E., DeBar, L.L., Kelleher, C., Lynch, F., Nunley, S., 2002. Overcoming depression on the internet (ODIN): a randomized controlled trial of an internet depression skills intervention program. J. Med. Internet Res. 4 (3).

Clarke, G., Eubanks, D., Kelleher, C., O'Connor, E., DeBar, L.L., Lynch, F., Nunley, S., Gullion, C., 2005. Overcoming depression on the internet (ODIN) (2): a

randomized trial of a self-help depression skills program with reminders. J. Med. Internet Res. 7 (2).
Colby, K., 1995. A computer program using cognitive therapy to treat depressed patients. Psychiatr. Serv. 46, 1223–1225.
Davis, T.C., Long, S.W., Jackson, R.H., Mayeaux, E.J., George, R.B., Murphy, P.W., et al., 1993. Rapid estimate of adult literacy in medicine: a shortened screening instrument. Fam. Med. 25 (6), 391–395.
Doherty, G., Coyle, D., 2009. Clinical evaluations and collaborative design: developing new technologies for mental health interventions. In: Proceedings of CHI 2009, Boston, MA.
Fann, J., Berry, D., Wolpin, S., Austin-Seymour, M., Bush, N., Halpenny, B., et al., 2009. Depression screening using the patient health questionnaire-9 administered on a touch screen computer. Psychooncology 18 (1), 14–22.
Francis, L., Weiss, B., Senf, J., Heist, K., Hargraves, R., 2007. Does literacy education improve symptoms of depression and self-efficacy in individuals with low literacy and depressive symptoms? A preliminary investigation. J. Am. Board Fam. Med. 20 (1), 23–27.
Garrity, T., 1981. Medical compliance and the clinician-patient relationship: a review. Soc. Sci. Med. 15E, 215–222.
Gazmararian, J., Baker, D., Parker, R., Blazer, D., 2000. A multivariate analysis of factors associated with depression: evaluating the role of health literacy as a potential contributor. Arch. Intern. Med. 160 (21), 3307–3314.
Ghanem, K., Hutton, H., Zenilman, J., Zimba, R., Erbelding, E., 2005. Audio computer assisted self interview and face to face interview modes in assessing response bias among STD clinic patients. Sex. Transm. Infect. 81 (5), 421–425.
Goodman, J., 2009. Women's attitudes, preferences and perceived barriers to treatment for perinatal depression. Birth 36 (1), 60–69.
Hasin, D., Goodwin, R., Stinson, F., Grant, B., 2005. Epidemiology of major depressive disorder: results from the national epidemiologic survey on alcoholism and related conditions. Arch. Gen. Psychiatry 62 (10), 1097–1106.
Hill, W., Weinert, C., Cudney, S., 2006. Influence of a computer intervention on the psychological status of chronically ill rural women: preliminary results. Nurs. Res. 55 (1), 34–42.
Horvath, A., Greenberg, L., 1989. Development and validation of the working alliance inventory. J. Couns. Psychol. 36 (2), 223–233.
Houston, T.K., Cooper, L.A., Vu, H.T., Kahn, J., Toser, J., Ford, D.E., 2001. Screening the public for depression through the internet. Psychiatr. Serv. 52, 362–367.
Jack, B.W., Chetty, V.K., Anthony, D., Greenwald, J.L., Burniske, G.M., Johnson, A.E., Forsythe, S.R., O'Donnell, J.K., Paasche-Orlow, M.K., Manasseh, C., Martin, S., Culpepper, L., 2009. The re-engineered hospital discharge program to decrease rehospitalization: a randomized, controlled trial. Ann. Intern. Med. 150, 178–187.
Kissinger, P., Rice, J., Farley, T., Trim, S., Jewitt, K., Margavio, V., et al., 1999. Application of computer-assisted interviews to sexual behavior research. Am. J. Epidemiol. 149 (10), 950–954.
Klein, B., Richards, J., Austin, D., 2006. Efficacy of internet therapy for panic disorder. J. Behav. Ther. Exp. Psychiatry 37 (3), 213–238.
Kroenke, K., Spitzer, R., Williams, J., 2001. The PHQ-9: validity of a brief depression severity measure. J. Gen. Intern. Med. 16, 606–613.
Krupnick, J., Sotsky, S., Elkin, I., Simmens, S., Moyer, J., Watkins, J., et al., 2006. The role of the therapeutic alliance in psychotherapy and pharmacotherapy outcome: findings in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. J. Consult. Clin. Psychol. 64, 532–539.
Lai, T., Larson, E., Rockoff, M., Bakken, S., 2008. User acceptance of HIV TIDES—tailored interventions for management of depressive symptoms in persons living with HIV/AIDS. J. Am. Med. Inform. Assoc. 15, 217–226.
Lincoln, A., Paasche-Orlow, M., Cheng, D., Lloyd-Travaglini, C., Caruso, C., Saitz, R., et al., 2006. Impact of health literacy on depressive symptoms and mental health-related: quality of life among adults with addiction. J. Gen. Intern. Med. 21 (8), 818–822.
Lincoln, A., Espejo, D., Johnson, P., Paasche-Orlow, M., Speckman, J., Webber, T., et al., 2008. Limited literacy and psychiatric disorders among users of an urban safety-net hospital's mental health outpatient clinic. J. Nerv. Ment. Dis. 196 (9), 687–693.
Lindau, S., Basu, A., Leitsch, S., 2006. Health literacy as a predictor of follow-up after an abnormal pap smear: a prospective study. J. Gen. Intern. Med. 21 (8), 829–834.
Lisetti, C., 2008. Embodied computer agents for psychotherapy. Paper presented at the CHI 2008 Workshop on Technology in Mental Health.
Mancuso, C., Rincon, M., 2006. Impact of health literacy on longitudinal asthma outcomes. J. Gen. Intern. Med. 21 (8), 813–817.
Marks, I.M., Mataix-Cols, D., Kenwright, M., Cameron, R., Hirsch, S., Gega, L., 2003. Pragmatic evaluation of computer-aided self-help for anxiety and depression. Br. J. Psychiatry 183, 57–65.
Morris, N., MacLean, C., Littenberg, B., 2006. Literacy and health outcomes: a cross-sectional study in 1002 adults with diabetes. BMC Fam. Pract. 7, 49.
Newman, J., Des Jarlais, D., Turner, C., Gribble, J., Cooley, P., Paone, D., 2002. The differential effects of face-to-face and computer interview modes. Am. J. Public Health 92 (2), 294–297.
Paasche-Orlow, M.K., Parker, R.M., Gazmararian, J.A., Nielsen-Bohlman, L.T., Rudd, R.R., 2005. The prevalence of limited health literacy. J. Gen. Intern. Med. 20 (2), 175–184.
Parashar, S., Rumsfeld, J., Spertus, J., Reid, K., Wenger, N., Krumholz, H., et al., 2006. Time course of depression and outcome of myocardial infarction. Arch. Intern. Med. 166, 2035–2043.
Pratt, L., Brody, D., 2008. Depression in the United States household population. NCHS Data Brief (7), 1–8.
Scherer, M., Herrmann-Lingen, C., 2009. Single item on positive affect is associated with 1-year survival in consecutive medical inpatients. J. Gen. Hosp. Psych. 31, 8–13.
Sirey, J., Bruce, M., Alexopoulos, G., Perlick, D., Friedman, S., Meyers, B., 2001. Stigma as a barrier to recovery: perceived stigma and patient-rated severity of illness as predictors of antidepressant drug adherence. Psychiatr. Serv. 52 (12), 1615–1620.
Slack, W., 2000. Patient-computer dialogue: a review. In: Yearbook of Medical Informatics. pp. 71–78.
Sourial, S., 1997. An analysis of caring. J. Adv. Nurs. 26, 1189–1192.
Sudore, R., Yaffe, K., Satterfield, S., Harris, T., Mehta, K., Simonsick, E., et al., 2006. Limited literacy and mortality in the elderly: the health, aging, and body composition study. J. Gen. Intern. Med. 21 (8), 806–812.
US Department of Health and Human Services, 2004. The President's New Freedom Commission on Mental Health, Subcommittee on Rural Issues: Background paper. DHHS, Rockville, Maryland.
Weinert, C., Cudney, S., Hill, W., 2008. Rural women, technology, and self-management of chronic illness. Can. J. Nurs. Res. 40 (3), 114–134.
Weiss, B., Francis, L., Senf, J., Heist, K., Hargraves, R., 2006. Literacy education as treatment for depression in patients with limited literacy and depression: a randomized controlled trial. J. Gen. Intern. Med. 21 (8), 823–828.
Weizenbaum, J., 1966. Eliza – a computer program for the study of natural language communication between man and machine. Commun. ACM 9 (1), 36–45.
Williams, M., Parker, R., Baker, D., Parikh, N., Pitkin, K., Coates, W., et al., 1995. Inadequate functional health literacy among patients at two public hospitals. JAMA 274 (21), 1677–1720.
World Health Organization, 2001. The World Health Report 2001 – Mental Health: New Understanding. World Health Organization, New Hope, Geneva, Switzerland.
Wright, J., 2004. Computer-assisted cognitive-behavior therapy. In: Wright, J. (Ed.), Review of Psychiatry: Cognitive-Behavior Therapy, vol. 23. American Psychiatric Publishing, Washington, DC, pp. 55–82.
Wright, J., Wright, A., Albano, A., Basco, M., Goldsmith, L., Raffield, M., et al., 2005. Computer-assisted cognitive therapy for depression: maintaining efficacy while reducing therapist time. Am. J. Psychiatry 162, 1158–1164.

Journal of Health Communication, 15:197–210, 2010

ISSN: 1081-0730 print/1087-0415 online
DOI: 10.1080/10810730.2010.499991

Routledge
Taylor & Francis Group

Usability of Conversational Agents by Patients with Inadequate Health Literacy: Evidence from Two Clinical Trials

TIMOTHY W. BICKMORE, LAURA M. PFEIFER, AND DONNA BYRON

Northeastern University College of Computer and Information Science, Boston, Massachusetts, USA

SHAULA FORSYTHE

Department of Family Medicine, Boston Medical Center/Boston University School of Medicine, Boston, Massachusetts, USA

LORI E. HENAULT

Section of General Internal Medicine, Boston Medical Center/Boston University School of Medicine, Boston, Massachusetts, USA

BRIAN W. JACK

Department of Family Medicine, Boston Medical Center/Boston University School of Medicine, Boston, Massachusetts, USA

REBECCA SILLIMAN

Section of Geriatrics, Boston Medical Center/Boston University School of Medicine, Boston, Massachusetts, USA

MICHAEL K. PAASCHE-ORLOW

Section of General Internal Medicine, Boston Medical Center/Boston University School of Medicine, Boston, Massachusetts, USA

Embodied Conversational Agents (ECA) are computer-animated characters that simulate face-to-face conversation with patients. These agents can be programmed with best practices in human-human health communication and used for automated health education and behavior change counseling interventions. Evidence is presented from two ongoing clinical trials demonstrating that patients at different levels of health literacy find these agents acceptable and easy to use for automated health communication interventions. Innovative computer interface systems can be used to ensure that inadequate health literacy not serve as a barrier to interventions using health information technology.

Address correspondence to Timothy W. Bickmore, Northeastern University, Computer Science, WVH202, 360 Huntington Ave, Boston, MA 02115, USA. E-mail: bickmore@ccs.neu.edu

There has been an explosion of interest and creativity in the field of health information technology, driven not only by the tremendous advantages of electronic medical records, but also by the great prospect for this technology to directly support patients for self-care and health behavior change. Research in this area has accelerated over the course of the past 25 years; however, the potential health benefits of this technology have not been realized. Two of the chief reasons for this are related to accessibility and usability. If patients cannot acquire the technologies or use them correctly, there is little possibility that such advances could lead to improvement in clinical outcomes.

Indeed, it is likely that current advances in patient-facing health information technology will exacerbate health disparities, as the benefits of such technologies will disproportionately accrue to the wealthiest, most educated, and technologically advanced members of society (Norman & Skinner, 2006; Bodie & Dutta, 2008). In particular, patients with inadequate health literacy are likely to be particularly vulnerable in this regard. People with inadequate health literacy are much less likely to use computers and have difficulty processing health information (Kutner, Greenberg, Jin, & Paulsen, 2006). Addressing disparities in access and usability is thus an essential element of addressing health disparities in general.

We have developed a computer interface—called an Embodied Conversational Agent (ECA)—that is usable by people with inadequate health literacy (Bickmore, Pfeifer, & Paasche-Orlow, 2009). The interface uses the universal and familiar format of face-to-face conversation, not just as an interface metaphor, but as the actual model of interaction. This is accomplished through the use of an animated character that talks to patients using synthetic speech and synchronized conversational nonverbal behavior, such as hand gestures, head nods, and eyebrow raises (Figure 1) (Cassell, Sullivan, Prevost, & Churchill, 2000). Patients talk to the character using touch screen input.

Figure 1. Embodied conversational agent interface for walking promotion trial.

Motivation for Using Embodied Conversational Agents with Inadequate Health Literacy Patients

Evidence suggests that face-to-face encounters with a health provider—in conjunction with written instructions—remain one of the best methods for communicating health information to patients in general, but especially those with inadequate health literacy (Qualls, Harris, & Rogers, 2002; Colcher & Bass, 1972; Madden, 1973; Morris & Halperin, 1979; Clark & Brennan, 1991). Face-to-face consultation is effective because it requires that the provider focus on the most salient information to be conveyed and that the information be delivered in a simple, conversational speaking style (Qualls et al., 2002). Protocols for "grounding" in face-to-face conversation—the use of verbal and nonverbal cues such as head-nods, gaze and acknowledgment tokens ("uh-huh," "OK") to signal mutual understanding (Clark & Brennan, 1991)—allows providers to dynamically assess a patient's level of understanding and repeat or elaborate information as necessary. Face-to-face conversation also allows providers to make their communication more explicitly interactive by asking patients to do, write, say, or show something that demonstrates their understanding (Doak, Doak, & Root, 1996). Finally, face-to-face interaction allows providers to use verbal and nonverbal behaviors, such as empathy (Frankel, 1995) and immediacy (Argyle, 1988; Richmond & McCroskey, 1995) to elicit patient trust, enabling better communication and satisfaction.

Given the efficacy of face-to-face consultation, Embodied Conversational Agents (ECA) show particular promise for conveying health information to patients with inadequate health literacy by simulating face-to-face conversation with a provider. These systems can produce verbal and nonverbal conversational behaviors that signify understanding, mark significance, and convey information in redundant channels of information (including speech intonation, hand gesture, facial display, body

Figure 2. Embodied conversational agent interface in rehospitalization trial (photo Glenn Kulbako).

posture shift, and eye gaze), to maximize message comprehension. They can use the verbal and nonverbal communicative behaviors used by health providers to establish trust and rapport with their patients in order to increase satisfaction and adherence to treatment regimens (Bickmore, Gruber, & Picard, 2005). They can adapt their messages to the particular needs of patients and to the immediate context of the conversation, since each utterance by the agent is dynamically composed (not just pre-recorded). They can emulate clinicians' extensive use of pointing gestures when explaining written materials to patients in order to clarify references and describe the structure and layout of the text (as in Figure 2) (Bickmore, Pfeifer, & Yin, 2008). Finally, they can provide health information in a consistent manner and in a low-pressure environment in which patients are free to take as much time as they need to thoroughly understand it.

ECA-Based Health Intervention Clinical Trials

We are currently using the ECA interface in two randomized clinical trials that specifically examine the role of health literacy. In one of these trials, the ECA is being used to teach patients being discharged from the hospital about their after hospital care plan. In the other it is being used to promote walking in older adults. The goal of the current analysis is to evaluate data from these ongoing trials regarding the usability of the ECA system for people with inadequate health literacy.

The two clinical trials of ECA-based health interfaces are being conducted at Boston Medical Center, a large urban safety-net hospital and ambulatory care center with an ethnically diverse patient population. Both studies use an ECA-based computer interface to communicate health information to patients, modeling best practices in health communication for patients with inadequate health literacy.

In the current analysis we are focused on measures related to satisfaction, usability and other process measures (in both studies we are blinded to health outcomes until trial completion: 30-day hospital utilization in the rehospitalization study and walking steps and fitness in the walking study).

Rehospitalization Trial

The first trial, entitled, "A RCT to Reduce Cardiopulmonary Rehospitalization" (PI: Jack, R01HL081307) is a two-armed intervention trial to improve patient education and safety in the transition between hospital and home with a primary goal of reducing 30-day hospital readmission. The system is designed to be used by patients in their hospital beds. The agent is deployed on a wheeled kiosk with a touch screen display attached to an articulated arm that can be positioned in front of patients (Figure 2). The system is designed to interact with patients once each day they are in the hospital, but the primary interaction is just before hospital discharge (75% of patients only had this final, discharge interaction, due to short hospital stays or logistical constraints). The final interaction is performed after the final list of discharge medications are approved ("medication reconciliation"), and typically just before the patient leaves the hospital. In this interaction, patients spend approximately half an hour using the system, to review the layout and contents of a personalized "After Hospital Care Plan" booklet that is produced for them and contains their post-discharge self-care instructions. The paper booklet is given to patients before their conversation with the agent, and the agent displays and reviews a digital version of the patient's booklet in the interface, so that patients can follow along

with the agent's explanation in their paper booklets to review medications, exercise and diet recommendations, and follow-up appointments. The specific approach to discharge education used in this project was modeled on our prior intervention—the Re-Engineered Hospital Discharge (RED)—which was delivered by a nurse (Jack, Chetty, & Anthony, 2009).

Rehospitalization Trial Methods

Participants

Participants in the rehospitalization study were English-speaking patients, 18 years of age or older, admitted to the teaching service of Boston Medical Center between October, 2008 and August, 2009. Patients were required to have a telephone, be able to comprehend study details and the consent process in English, and have plans to be discharged to a U.S. community. Patients were not enrolled if they were admitted from a skilled nursing facility or other hospital, admitted for a planned hospitalization, on hospital precautions, on suicide watch, deaf, or blind. Of the 417 participants enrolled to date into the parent study (of a planned 750), 208 were randomized into the ECA intervention arm of the study. Of these, 143 completed all measures necessary for our analyses (there were no significant differences in demographic characteristics between those who completed all measures and those who left the hospital prior to completing the study protocol).

Measures

Health Literacy

Health literacy was assessed using the Rapid Estimate of Adult Literacy in Medicine (REALM) (Davis, Long, & Jackson, 1993). The sample was split into "adequate" and "inadequate" health literacy groups, using a REALM score of 9th grade and above, as other authors have done (Lindau, Basu, & Leitsch, 2006; Mancuso & Rincon, 2006; Sudore, Yaffe, & Satterfield, 2006; Lincoln, Paasche-Orlow, & Cheng, 2006).

Usability

Usability was assessed using single scale-measure self-report items to assess overall satisfaction with the ECA, ease of use, desire to continue working with the ECA, and preference for the ECA relative to human health providers, in addition to session duration.

Attitudes Towards the Agent—Therapeutic Alliance

Patient attitude towards the ECA was assessed using a measure of therapeutic alliance, specifically the affective bond subscale of the Working Alliance Inventory. This is a self-reported 12-item Likert scale questionnaire assessing the emotional dimension of a patient's trust and belief that they can work together with a provider to achieve desired therapeutic outcomes (Horvath & Greenburg, 1989).

Procedure

Participants were provided with brief training on how to "talk" to the ECA, in which the agent walks on the screen and greets the participant; participants are then told to

"touch what you want to say on the screen" (that is typically the extent of the training). At the conclusion of their interaction with the ECA they answered questions regarding usability and attitudes toward the gent just prior to leaving the hospital. All self-report measures were verbally collected by research staff to accommodate patients with limited literacy.

Rehospitalization Trial Results

Demographics and Health Literacy

Table 1 shows demographics of the study population. Participants with inadequate health literacy in the rehospitalization study were significantly older, less educated, and more likely to be non-White compared with participants in that study with adequate health literacy. Participants with inadequate health literacy also had significantly lower levels of computer literacy compared to participants with adequate health literacy.

Usability

Participants reported very high levels of overall satisfaction and ease of use, regardless of health literacy level: 78% of all participants scored satisfaction a 7 on a

Table 1. Subject demographics by health literacy level

Health literacy level	Inadequate	Adequate	*p* value
Rehospitalization Study			
N	68	75	
Sex (% Male)	56.9	50.0	n.s.
Age (range 20–84)	52.7	46.6	.004
Race: % African American	63.9	48.6	
Race: % White	12.5	35.1	<.05
Race: % Other	23.6	16.3	
% Hispanic or Latino	13.9	10.8	
Highest grade completed	11.6	12.9	.002
Computer Literacy (1 = never use one; 4 = expert)	2.01	2.73	<.001
Walking Study			
N	15	18	
Sex (% Male)	33.3	22.2	n.s.
Age (range 65–85)	73.0	73.5	n.s.
Race: % African American	86.7	61.1	
Race: % White	6.7	22.2	n.s.
Race: % Other	6.6	16.7	
% Hispanic or Latino	20.0	0.0	
Highest grade completed	11.3	13.3	n.s.
Computer Literacy	1.4	1.8	.08

Health literacy assessed via REALM for Rehospitalization Study and TOFHLA for Walking Study.

7-point Likert-type scale (with 7 = "very satisfied"), and 78% scored ease of use a 1 on a 7-point scale (with 1 = "very easy to use"). In addition, participants with inadequate health literacy showed a trend of greater satisfaction with the ECA compared to participants in that study with adequate health literacy.

None of the other usability measures were significantly different across health literacy levels.

Attitudes Toward the Agent

Participants scored well above the Likert scale midpoint on overall mean Working Alliance Bond subscale scores, regardless of health literacy level, and only 11% of participants scored below the midpoint of the composite measure. In addition, there were no significant differences between literacy groups on overall Working Alliance scores. However, differences on a few of the individual items in the scale reached significance, indicating a greater degree of personification of the agent (mutual respect, importance of relationship with the agent) by participants with inadequate health literacy.

Geriatrics Walking Promotion Trial

The second trial, entitled, "Computer Agents to Promote Walking in Older Adults with Low Health Literacy" (PI: Silliman, R01AG028668) is a two-armed intervention trial to promote walking in older adults with a primary goal of improving the number of steps per day at 12 months. Older adult ambulatory clinic patients at Boston Medical Center are given pedometers which link to tablet-PC computers. Intervention participants are given a tablet-PC to use at home for 2 months and are asked to interact with the ECA daily to set and discuss walking goals (Figure 1). In addition, participants can interact with the agent on a kiosk in the waiting room of their primary care provider.

Walking Trial Methods

Participants

Participants in the walking study were English-speaking patients, 65 years or older, who attend the geriatrics or internal medicine ambulatory care clinics at Boston Medical Center between April, 2009 and September, 2009. Patients were required to speak and read English at a level required to interact with the ECA (via a screening conversation with the agent) and to understand the study protocol, be inactive but medically able to begin a moderate intensity physical activity program, and free of cognitive impairment and significant depressive symptoms. Of the 88 participants enrolled to date into the parent study (of a planned 270), 44 were randomized into the ECA intervention arm of the study and 2-month study measures were obtained from 33 of these.

Measures

Health Literacy

Health literacy was assessed using the Test of Functional Health Literacy in Adults (TOFHLA) (Parker, Baker, Williams, & Nurss, 1995). A different measure was used relative to the rehospitialization trial due to the different patient populations and

study settings. Patients with subclinical dementia can often pronounce a word correctly but not know what the word means, invalidating REALM results. As this is more likely to occur in older cohorts, we chose to avoid the REALM in the walking study, and used the TOFHLA. However, as the TOFHLA takes more time to administer, it was not the best choice for rushed hospital environments, especially those with relatively younger adult populations. Both of these measures reflect print literacy and reading ability (Berkman, Pignone, Sheridan, & Lohr, 1994) and so may not be the most accurate assessments of ability to act on health information communicated verbally.

The sample was split into adequate and inadequate health literacy groups, using a TOFHLA score of 23 or above, as other authors have done (Lindau et al., 2006; Mancuso & Rincon, 2006; Sudore et al., 2006; Lincoln et al., 2006).

Usability

Usability was assessed through actual voluntary use of the system during the first two months in which patients had the tablet computer at home, based on the tablet log files. Measures included the number of sessions completed out of 60 possible daily conversations, the average duration of each session, and the percent of sessions in which participants plugged in their pedometer (the agent asked them to plug it in every session).

Attitudes Towards the Agent—Therapeutic Alliance

Attitude towards the agent was assessed using the affective bond subscale of the Working Alliance Inventory, as in the rehospitalization trial (Horvath & Greenberg, 1989).

Procedure

Participants were provided with the same brief ECA training as in the rehospitalization study, given at time of enrollment, before being sent home with a tablet computer for two months of home-based interactions with the agent. Assessments of attitudes toward the ECA were administered at an in-person research interview immediately following these 2 months. All self-report measures were verbally collected by research staff to accommodate patients with limited literacy.

Analysis

In order to examine the trends in participant use of the system over time, we analyzed the sessions data using mixed-effect modeling. All analysis was performed using R 2.9.0 (R Development Core Team, 2008) with the "nlme" package, fitting linear mixed-effect regression models to the sessions per week and literacy category data. Best fit results were for a model with random effects for intercept but not study week (slope).

Walking Trial Results

Demographics and Health Literacy

As in the rehospitalization study, participants with inadequate health literacy had lower levels of computer literacy compared to participants with adequate health literacy, although this difference was only trending towards significance, likely due to the smaller sample size (Table 1).

Table 2. Outcomes by health literacy level

Health literacy level	Inadequate	Adequate	*p* value
Rehospitalization Study			
Satisfaction (1 = not at all; 7 = very much)	6.57	6.45	.083
Ease of Use (1 = very easy; 7 = very difficult)	1.82	1.83	n.s.
Desire to Continue with Agent (1 = not at all; 7 = very much)	5.82	5.39	n.s.
Prefer Human Provider over Agent (1 = definitely prefer doctor or nurse; 7 = definitely prefer agent)	4.50	4.12	n.s.
Average session time (minutes)	31.62	27.38	n.s.
WAI* Bond (overall composite)	5.80	5.49	n.s.
I am comfortable with the agent.	5.58	5.78	n.s.
The agent and I understand each other.	5.67	5.68	n.s.
The agent likes me.	5.50	5.29	n.s.
The agent is concerned about my welfare.	6.16	5.64	n.s.
The agent and I respect each other.	6.24	5.59	.027
The agent is honest about her feelings towards me.	4.83	5.29	n.s.
I am confident in the agent's ability to help me.	6.43	6.20	n.s.
The agent appreciates me.	5.97	5.52	n.s.
The agent and I trust one another.	5.68	5.32	n.s.
My relationship with the agent is important to me.	5.82	4.99	.012
The agent cares about me, even if I do something wrong.	5.24	4.74	n.s.
The agent will keep working with me, even if I say something wrong.	5.76	5.81	n.s.
Walking Study			
Sessions completed (of 60 possible)	26.73	38.39	.078
Average time per session (minutes)	7.49	7.67	n.s.
Sessions with pedometer uploads (percent)	64.00	83.55	.058
WAI* Bond (overall composite)	5.71	5.24	n.s.
I am comfortable with the agent.	5.67	4.94	n.s.
The agent and I understand each other.	6.20	4.83	.015
The agent likes me.	5.93	5.93	n.s.
The agent is concerned about my welfare.	5.93	5.39	n.s.
The agent and I respect each other.	6.20	5.28	n.s.
The agent is honest about her feelings towards me.	5.60	5.11	n.s.
I am confident in the agent's ability to help me.	6.20	5.50	n.s.
The agent appreciates me.	5.67	5.22	n.s.

(*Continued*)

Table 2. Continued

Health literacy level	Inadequate	Adequate	*p* value
The agent and I trust one another.	5.60	5.22	n.s.
My relationship with the agent is important to me.	5.60	5.06	n.s.
The agent cares about me, even if I do something wrong.	5.93	5.11	n.s.
The agent will keep working with me, even if I say something wrong.	3.73	5.61	.011

All t-tests except Satisfaction (Mann-Whitney due to ceiling effect).
*WAI: Working Alliance Inventory (all items Likert scale, 1 = disagree completely; 7 = agree completely).

Usability

Mixed effect regression indicates that participants with inadequate health literacy completed fewer home-based conversations with the ECA compared to participants with adequate health literacy ($p < .05$), Note that a simple *t*-test on total number of sessions also shows this result (approaching significance, Table 2). Regression results also indicate a significant decrease in home-based conversations with the ECA over time for all participants of -0.29 sessions/week ($p < .001$). There was no significant interaction between sessions per week and literacy category; participants had similar patterns of decreasing use over time, regardless of literacy category.

There were no significant differences in session durations between literacy categories, but there was a trend for participants with adequate health literacy to plug in their pedometers more frequently compared to participants with inadequate health literacy ($p = .058$, Table 2).

Attitudes Toward the Agent

As in the rehospitalization study, participants scored well above the Likert scale midpoint on overall mean Working Alliance Bond subscale scores, regardless of health literacy level, and only 9% of participants scored below the midpoint of the composite measure. In addition, there were no significant differences between literacy groups on overall Working Alliance scores. However, differences on a few of the individual items in the scale reached significance, indicating a greater degree of personification of the agent (mutual understanding) and a lower level of understanding of the technology (thinking the agent would discontinue use if the participant said the wrong thing) by participants with inadequate health literacy.

Discussion

Overall, there were very few differences in measures of acceptance and usability between patients with adequate and inadequate health literacy, suggesting that ECAs are approachable and usable by patients regardless of health literacy level. In the few measures in which there were significant or near-significant differences on health literacy, these were mostly in favor of patients with inadequate health literacy,

indicating that ECAs may be even more acceptable to this population than to patients with adequate health literacy.

In addition to the theoretical reasons why ECAs may be ideal interfaces for patients with inadequate health literacy, described earlier, patients interviewed in the pilot studies that preceded the two trials provided a better understanding of their reasons for accepting the technology (Bickmore, Pfeifer, & Jack, 2009; Bickmore, Caruso, Clough-Gorr, & Heeren, 2005). Patients in both pilots indicated that the system was very easy to use, even if they had little or no experience with computers:

- "*I don't like computers but that was easy*." (rehospitalization pilot)
- "*That is so easy. That is so good. Regular computers I don't do. But, that was so easy, even a baby could do that*." (walking pilot)

Patients in the rehospitalization pilot indicated that they liked being able to take as much time as they needed to understand everything, something they said that their doctors or nurses typically did not provide:

- "*Sometimes doctors just talk and assume you understand what they're saying. With a computer you can go slow, go over things again and she checks that you understand*."
- "*I prefer Louise [the name of the ECA character], she's better than a doctor, she explains more, and doctors are always in a hurry*."

Patients in both pilots were also mostly positive about the interventions:

- "*It was the best thing that happened to me, to have something that pushed me out and get me walking*." (walking pilot)
- "*She's nice. She's really good. Really good. She asks you the right questions. She tells if you if you're not doing up to par, you know, and all that. And if you're doing good, she'll tell you. If you're not she'll tell you. And it's honest. And it works. It really does. I like it. I like talking to her*." (walking pilot)
- "*She treated me like a real person! She's not like a computer. This is awesome work! This is really excellent*." (rehospitalization pilot)
- "*I've had problems with, not this hospital, but other hospitals. I wasn't given the quality time that this lady gave me*." (rehospitalization pilot)

One area of possible concern is that patients with adequate health literacy in the walking trial completed more sessions with the ECA compared to patients with inadequate health literacy. This may indicate that, despite having similar attitudes towards and satisfaction with the agent and despite finding the system easy to use, there may be other important factors such as patient activation that dictate the amount of use. However, the relationship between intervention dose and health outcomes in behavioral studies can be complex, and it could even be that fewer sessions result in better outcomes. The ECA provides an accessible and usable communication channel for patients irrespective of health literacy, but more research is required to ascertain contexts in which dose is important and then to tailor information and counseling dialogue content to ensure that a given intervention is effective for patients with inadequate health literacy.

Patients with inadequate health literacy appear to anthropomorphize ECAs more than patients with adequate health literacy, as reflected by specific items related to mutual understanding and respect, and belief that the agent may decide to stop working with them if they say something wrong. Although this indicates a

general lack of understanding of the underlying technology, it may ultimately prove beneficial for these patients if the increased personification leads to a greater sense of working alliance and increased adherence to the ECA's recommendations. Some patients may actually confuse the agent with a real person (e.g., if delirious in the hospital), which could be partially addressed by having both the humans administering the agent and the agent itself periodically remind users that it is just a computer. Another concern is that the results may indicate that patients with adequate health literacy do not like the social aspects of the interactions, feeling that they are unnecessary, slow, or even disingenuous. Future systems may allow patients to choose more conventional graphical user interfaces that let them work through the information in a session more efficiently.

Future Work

Our immediate future plans are to complete the rehospitalization and walking trials in order to demonstrate efficacy—in terms of clinically important health outcomes—regardless of health literacy level.

Now that we have established that ECAs can provide an acceptable and usable health communication channel for patients with inadequate health literacy, the opportunities for developing patient and consumer education and counseling interventions are limitless. Specific areas that we are investigating include:

- Automated explanation of written medical information to patients with varying levels of health literacy (Bickmore, Pfeifer, & Paasche-Orlow, 2009).
- Linguistically and culturally tailored health interventions, such as exercise promotion for older bilingual Latino adults (Yin, Bickmore, Byron, & Cortes, 2010).
- Longitudinal health behavior change interventions, in which alliance with the ECA is used to promote retention in the intervention as well as adherence (Bickmore, Schulman, & Yin, 2010).
- Deployment on other computer platforms, including mobile devices (Bickmore & Mauer, 2009).

A final important area of ongoing research is the automatic adaptation of the computer interface in response to patient characteristics and needs. Our finding that patients with high levels of computer literacy are less satisfied with the ECA may indicate that such patients should be given the option of using a more traditional computer interface to more efficiently access the information they need, while patients with low computer and/or health literacy would use the ECA. In addition, in some of our studies we have found that nurses provide different information to patients depending on their level of health literacy—providing more technical detail to patients with adequate health literacy, but providing more scaffolding (information about document structure) to patients with inadequate health literacy (Bickmore, Pfeifer, & Yin, 2008)—and this difference in presentation could also be emulated by an ECA that dynamically adjusts its dialogue based on patient needs.

References

Argyle, M. (2008). *Bodily communication*. New York: Methuen & Co. Ltd.

Berkman, N., Pignone, M., Sheridan, S., & Lohr, K. (2004). Literacy and Health Outcomes. Evidence Report/Technology Assessment No. 87 University of North Carolina Evidence-based Practice Center.

Bickmore, T., Caruso, L., Clough-Gorr, K., & Heeren, T. (2005). It's just like you talk to a friend. *Relational Agents for Older Adults. Interacting with Computers*, *17*, 711–35.

Bickmore, T., Gruber, A., & Picard, R. (2005). Establishing the computer-patient working alliance in automated health behavior change interventions. *Patient Education and Counseling*, *59*, 21–30.

Bickmore, T., Pfeifer, L., & Yin, L. (2008). The role of gesture in document explanation by embodied conversational agents. *International Journal of Semantic Computing*, *2*, 47–70.

Bickmore, T., & Mauer, D. (2009). Context awareness in a handheld exercise agent. *Pervasive and Mobile Computing (Special Issue on Pervasive Health and Wellness)*, *5*, 226–235.

Bickmore, T., Pfeifer, L., & Jack, B. W. (2009). Taking the time to care: Empowering low health literacy hospital patients with virtual nurse agents. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*; 2009; Boston, MA.

Bickmore, T., Pfeifer, L., & Paasche-Orlow, M. (2009). Using computer agents to explain medical documents to patients with low health literacy. *Patient Education and Counseling*, *75*, 315–320.

Bickmore, T., Schulman, D., & Yin, L. (2010). Maintaining engagement in long-term interventions with relational agents. *International Journal of Applied Artificial Intelligence*, *24*(6), 648–666.

Bodie, G., & Dutta, M. (2008). Understanding health literacy for strategic health marketing: eHealth literacy, health disparities, and the digital divide. *Health Marketing Quarterly*, *25*, 175–203.

Cassell, J., Sullivan, J., Prevost, S., & Churchill, E. (Eds). (2000). *Embodied conversational agents*. Cambridge, MA: The MIT Press.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine., & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127–149). Washington, DC: American Psychological Association.

Clinite, J., & Kabat, H. (1976). Improving patient compliance. *Journal of the American Pharm Association*, *16*, 74–76.

Colcher, I., & Bass, J. (1972). Penicillin treatment of streptococcal pharyngitis. *Journal of the American Medical Association*, *222*, 657–659.

Davis, T., Long, S., & Jackson, R. (1993). Rapid estimate of adult literacy in medicine: A shortened screening instrument. *Family Medicine*, *25*, 391–395.

Doak, C., Doak, L., & Root, J. (1996). *Teaching patients with low literacy skills* (2nd ed.). Philadelphia, PA: J.B. Lippincott.

Frankel, R. (1995). Emotion and the physician-patient relationship. *Motivation and Emotion*, *19*, 163–173.

Horvath, A., & Greenberg, L. (1989). Development and Validation of the working alliance inventory. *Journal of Counseling Psychology*, *36*, 223–233.

Jack, B. W., Chetty, V. K., Anthony, D., et al. (2009). The re-engineered hospital discharge program to decrease rehospitalization: A randomized, controlled trial. *Annals of Internal Medicine*, *150*, 178–187.

Kutner, M., Greenberg, E., Jin, Y., & Paulsen, C. (2006). *The health literacy of America's adults: Results from the 2003 National Assessment of Adult Literacy*. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Lincoln, A., Paasche-Orlow, M., & Cheng, D., et al. (2006). Impact of health literacy on depressive symptoms and mental health-related: Quality of life among adults with addiction. *Journal of General Internal Medicine*, *21*, 818–822.

Lindau, S., Basu, A., & Leitsch, S. (2006). Health literacy as a predictor of follow-up after an abnormal Pap smear: A prospective study. *Journal of General Internal Medicine*, *21*, 829–834.

Madden, E. (1973). Evaluation of outpatient pharmacy patient counseling. *Journal of the American Pharm Association*, *13*, 437–443.

Mancuso, C., & Rincon, M. (2006). Impact of health literacy on longitudinal asthma outcomes. *Journal of General Internal Medicine*, *21*, 813–817.

Morris, L., & Halperin, J. (1979). Effects of written drug information on patient knowledge and compliance: A literature review. *American Journal of Public Health*, *69*, 47–52.

Norman, C., & Skinner, H. (2006). eHealth literacy: Essential skills for consumer health in a networked world. *Journal of Medical Internet Research*, *16*, 2.

Parker, R., Baker, D., Williams, M., & Nurss, J. (1995). The test of functional health literacy in adults (TOFHLA): A new instrument for measuring patients' literacy skills. *Journal of General Internal Medicine*, *10*, 537–541.

Qualls, C., Harris, J., & Rogers, W. (2002). Cognitive-linguistic aging: Considerations for home health care environments. In W. Rogers & A. Fisk (Eds.), *Human factors interventions for the health care of older adults* (pp. 47–67). Mahwah, NJ: Lawrence Erlbaum.

R Development Core Team. (2008). R: A language and environment for statistical computing. Available at http://www.R-project.org

Richmond, V., & McCroskey, J. (1995). Immediacy. In *Nonverbal behavior in interpersonal relations* (pp. 195–217). Boston: Allyn & Bacon.

Sudore, R., Yaffe, K., Satterfield, S., Harris, T. B., Mehta, K. M., Simonsick, E. M., et al. (2006). Limited literacy and mortality in the elderly: The health, aging, and body composition study. *Journal of General Internal Medicine*, *21*, 806–812.

Yin, L., Bickmore, T., Byron, D., & Cortes, D. (2010). Cultural and linguistic adaptation of relational agents for health counseling. In *Workshop on interactive systems in healthcare*. Atlanta, GA.

Blood et al. BMC Medical Research Methodology 2010, 10:16
http://www.biomedcentral.com/1471-2288/10/16

BMC Medical Research Methodology

RESEARCH ARTICLE Open Access

Performance of mixed effects models in the analysis of mediated longitudinal data

Emily A Blood[1,2*], Howard Cabral[1], Timothy Heeren[1], Debbie M Cheng[1]

Abstract

Background: Linear mixed effects models (LMMs) are a common approach for analyzing longitudinal data in a variety of settings. Although LMMs may be applied to complex data structures, such as settings where mediators are present, it is unclear whether they perform well relative to methods for mediational analyses such as structural equation models (SEMs), which have obvious appeal in such settings. For some researchers, SEMs may be more difficult than LMMs to implement, e.g. due to lack of training in the methodology or the need for specialized SEM software. It therefore is of interest to evaluate whether the LMM performs sufficiently in a scenario particularly suitable for SEMs. We focus on evaluation of the total effect (i.e. direct and indirect) of an exposure on an outcome of interest when a mediating factor is present. Our aim is to explore whether the LMM performs as well as the SEM in a setting that is conducive to using the SEM.

Methods: We simulated mediated longitudinal data from an SEM where a binary, main independent variable has both direct and indirect effects on a continuous outcome. We conducted analyses with both the LMM and SEM to evaluate the performance of the LMM in a setting where the SEM is expected to be preferable. Models were evaluated with respect to bias, coverage probability and power. Sample size, effect size and error distribution of the simulated data were varied.

Results: Both models performed well in a range of settings. Marginal increases in power estimates were observed for the SEM, although generally there were no major differences in performance. Power for both models was good with a sample of size of 250 and a small to medium effect size. Bias did not substantially increase for either model when data were generated from distributions that were both skewed and kurtotic.

Conclusions: In settings where the goal is to evaluate the overall effects, the LMM excluding mediating variables appears to have good performance with respect to power, bias and coverage probability relative to the SEM. The major benefit of SEMs is that it simultaneously and efficiently models both the direct and indirect effects of the mediation process.

Background

A common method of handling longitudinal data is through linear mixed effects models (LMMs) [1]. These models account for the correlation of observations and allow estimation of the effect of predictor variables on repeated outcomes. They are relatively easy to implement and their regression parameters have a clear interpretability.

Complex relationships often exist among the variables studied, however, and it may be of interest to explicitly model the hypothesized causal pathways between independent variables and outcomes. Although multiple mixed effects models can be fit to evaluate mediation (see e.g. Krull and MacKinnon [2] and Baron and Kenny [3]), methods for mediational analyses, such as Structural Equation Models (SEMs), are necessary to simultaneously model mediated relationships. However, when the primary aim of an analysis is to determine the total effect (i.e. direct and indirect) of an exposure on an outcome of interest, it is unclear what the impact of explicit modeling of the mediated relationship is on power, bias, and on coverage probability for the main research aim.

SEMs are a well known and commonly used data analysis technique in the social sciences, and is becoming increasingly popular in many clinical research areas. The

* Correspondence: emily.blood@childrens.harvard.edu
[1]Department of Biostatistics, Boston University School of Public Health, 715 Albany Street Crosstown Center-3rd Floor Boston, MA 02118, USA

SEM framework is a general modeling framework and allows the modeling of potentially complex relationships among observed and latent variables and can be applied in the longitudinal data setting.

There has been much previous work in applying SEMs to longitudinal data analyses [4-9], and the equivalence of LMMs and special cases of SEMs in settings without mediating variables has been well documented in the SEM literature [5,7,10-15]. The major advantages of SEMs are that they have the capability to incorporate measurement error on the variables in the model [5,8,13], allow explicit modeling of relationships involving mediating variables [16] and are able to decompose direct and indirect effects [11]. Disadvantages of SEMs that have been previously noted include potentially large sample size requirements and potential problems with skewed data. It is also essential for SEMs that investigators have clear hypotheses on the causal pathways between variables [9]. In addition, SEMs may be more difficult than LMM to implement, e.g. due to a lack of training in the methodology or the need for specialized SEM software. Given these potential limitations, it is of interest to explore whether a LMM performs well relative to an SEM in settings where mediation is present.

The purpose of this paper is to conduct a simulation study in a mediated longitudinal setting to evaluate whether a LMM performs sufficiently with respect to power, bias and coverage in a scenario that is conducive to using an SEM.

Methods

Setting

We consider a longitudinal setting similar to a study by Samet et al [17] evaluating the impact of heavy alcohol consumption on HIV disease progression. The data arise from a prospective cohort study in which the primary outcome, CD4 cell count, is assessed every 6 months for three years (i.e. 6 measures of CD4 count across time for each subject) and heavy alcohol consumption, the main independent variable, is assessed only at baseline. A potential mediator of the relationship between heavy alcohol consumption and HIV disease progression is adherence to antiretroviral therapy (ART) as it has been demonstrated that alcohol consumption may worsen a patient's ability to adhere to ART thereby leading to worse disease progression. In the current setting we assume that ART is assessed only at baseline. In addition to an indirect effect mediated by ART, alcohol consumption could also have a direct biological effect on CD4 cell count. The primary objective of the analysis is to evaluate the overall impact (direct plus indirect effect) of heavy alcohol use on CD4 cell count. Figure 1 shows a simple diagram illustrating the relationship between heavy alcohol consumption, ART adherence and the outcome CD4 cell count.

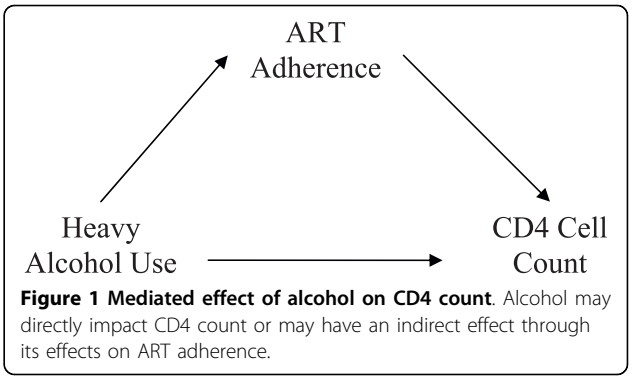


Figure 1 Mediated effect of alcohol on CD4 count. Alcohol may directly impact CD4 count or may have an indirect effect through its effects on ART adherence.

A standard analytic approach for analyzing these data would be to fit a LMM, which can account for correlation due to repeated assessments of CD4 cell count from the same subject and adjust for potential confounders. Alternatively, an SEM could be fit to the data which would explicitly model the hypothesized pathways between heavy alcohol consumption and CD4 count.

In this setting, where the main objective is to determine the total effect of heavy alcohol use on CD4 cell count, it is unclear whether a LMM can perform as well as an SEM, a method often preferred for analyzing mediated longitudinal data.

General SEM Formulation

There are two components to an SEM, the measurement model and the structural model. The measurement model relates unobserved latent variables and covariates to outcomes and exposure indicators. This model attempts to capture measurement error in observed variables. The structural model relates covariates and latent variables to latent variables. This model attempts to capture individual variation in the latent variables.

Using the same notation as Sanchez [15], the general model is expressed as:

Measurement Model

$$\begin{pmatrix} \mathbf{X_i} \\ \mathbf{Y_i} \end{pmatrix} = \nu + \Lambda \mathbf{U_i} + \mathbf{K} \mathbf{Z_i} + \epsilon_\mathbf{i} \quad (1)$$

Structural Model

$$\mathbf{U_i} = \alpha + \mathbf{B} \mathbf{U_i} + \mathbf{\Gamma} \mathbf{Z_i} + \zeta_\mathbf{i} \quad (2)$$

In the above equations, i indexes the individual, with $i = 1,..., N$ where N is the number of individuals. For the ith individual, $\mathbf{Y_i}$ is a vector of observed outcomes, $\mathbf{X_i}$ is a vector observed exposure indicators. $\mathbf{U_i}$ is a vector of latent variables and $\mathbf{Z_i}$ is a vector of observed, fixed

covariates. Although $\mathbf{U_i}$ appears on both sides of the matrix equation, the diagonal elements of **B** are zeros so that the same element of $\mathbf{U_i}$ would not appear on both the left- and right-hand side of a given equation. is a matrix of coefficients associated with $\mathbf{U_i}$, **K** is a matrix of coefficients associated with $\mathbf{Z_i}$ and ε_i is a vector of random residual errors for the measurement model. In the structural model, **B** is a matrix of coefficients (where the diagonal elements are zeros) associated with $\mathbf{U_i}$, Γ is a matrix of coefficients associated with $\mathbf{Z_i}$ and ζ_i is a vector of random residual errors for the structural part of the model. The mean of random residual errors for both the measurement and structural models are assumed to be zero. **Σ** is the covariance matrix of the residual errors of the measurement model (ε_i), and **Ψ** is the covariance matrix of the residual errors of the structural model (ζ_i). The $\mathbf{X_i}$ and $\mathbf{Y_i}$ are assumed to be multivariate normal (MVN). The errors in Equations 1 and 2 are assumed to be independent. Parameters are usually estimated via maximum likelihood, with the objective of minimizing the distance between the observed and model-based mean and covariance structure [16].

SEM Simulation Model

The SEM framework was used to generate the mediated longitudinal data for the simulation studies since the aim is to evaluate whether a LMM performs sufficiently in the setting where an SEM is presumed to be optimal. The scenario in which we simulated data is an extension of a specific SEM often referred to as a latent growth curve model or latent curve model [4,18]. In the latent growth curve model, the outcome variables are influenced by random intercept and slope variables. These variables are latent and can be influenced by predictors and other covariates. Let i index the individual (i = 1,..., N) and j index the time-point (j = 1,..., T), where T is the number of measurement times. In the current study, we considered a setting with a single continuous covariate (z_{1i}), such as age, and a single binary independent variable (z_{2i}) of primary interest, heavy alcohol consumption, predicting repeated observations of the outcome (Y_{ij}), CD4 cell count. Heavy alcohol use influences the outcome CD4 cell count through the random intercept and slope variables. In addition, ART adherence is a mediating variable (x_i) which influences CD4 count through the random intercept and slope variables. The variable ART adherence is said to be a mediator because the primary independent variable, heavy alcohol use, may affect CD4 count not only directly but also indirectly through ART adherence. We considered a setting with 6 time-points (T = 6) and illustrate the SEM model with the path diagram in Figure 2. Using the notation we have described above for SEMs and eliminating the subject index i for simplicity, our measurement and structural model for the scenario illustrated in Figure 2 can be written as:

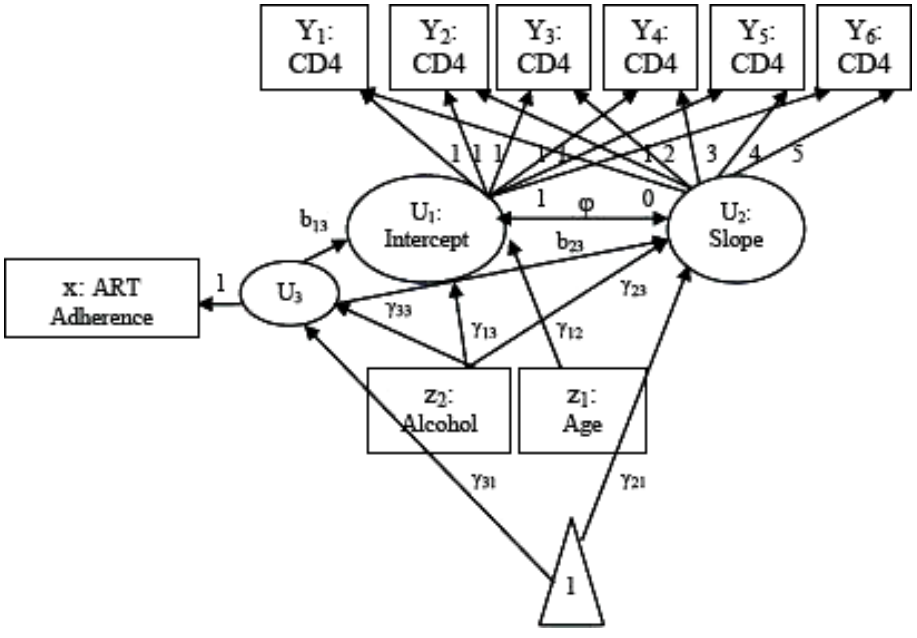


Figure 2 Path diagram for mediated longitudinal data. path diagram for mediated longitudinal data with the outcome (Y) measured at six occasions, a continuous covariate (z_1) measured once, a dichotomous main independent predictor (z_2) measured once variable and a continuous mediator (x) measured once.

Measurement Model

$$\begin{aligned} Y_j &= U_1 + t_j U_2 + \epsilon_j \\ x &= U_3 \end{aligned} \tag{3}$$

Structural Model

$$\begin{aligned} U_1 &= \gamma_{11} + \gamma_{12} z_1 + \gamma_{13} z_2 + b_{13} U_3 + \zeta_1 \\ U_2 &= \gamma_{21} + \gamma_{23} z_2 + b_{23} U_3 + \zeta_2 \\ U_3 &= \gamma_{31} + \gamma_{33} z_2 + \zeta_3 \end{aligned} \tag{4}$$

or in matrix notation as:

Measurement Model

$$\begin{aligned} \mathbf{Y_i} &= \Lambda \mathbf{U_i} + \epsilon_\mathbf{i} \\ x &= U_3 \end{aligned}$$

Structural Model

$$\mathbf{U_i} = \mathbf{B U_i} + \mathbf{\Gamma Z_i} + \zeta_\mathbf{i}$$

The latent intercept and slope are represented by U_1 and U_2, respectively. The continuous covariate (age) and main independent variable (heavy alcohol use) are represented by z_1 and z_2, respectively. Based on the model formulation presented in Equations 1 and 2, ART adherence (x) is considered both an outcome (as it is influenced by the main independent variable) and a predictor (as it influences the random intercept and slope), and therefore appears on the right- and left-hand side of the above equations. However, using the above formulation, only latent variables can be both outcomes and predictors. Thus to incorporate x as a mediator and stay within the framework defined by Sanchez [15], we must add an additional latent variable (U_3) to the model that is exactly equal to x; x can then be viewed as an indicator of this latent variable. Time is incorporated into the model by populating the Λ matrix from Equation 1 with the fixed times of measurements (t_j).

It can be shown that for a given outcome, Y_j, at time t_j, the predictive formula is:

$$\begin{aligned} Y_j &= b_{13}\gamma_{31} + b_{13}\gamma_{33} z_2 + \gamma_{12} z_1 + \gamma_{13} z_2 + \\ & b_{23}\gamma_{31} t_j + b_{23}\gamma_{33} z_2 t_j + \gamma_{23} z_2 t_j \\ & + (\zeta_1 + b_{13}\zeta_3) + (\zeta_2 + b_{23}\zeta_3) t_j + \epsilon_j \end{aligned} \tag{5}$$

The interpretation of the key model parameters of interest are as follows:

1. b_{13} is the effect of the mediating variable on the latent intercept.
2. b_{23} is the effect of the mediating variable on the latent slope.
3. γ_{33} is the effect of the main independent variable on the mediating variable.
4. γ_{23} is the effect of the main independent variable on the latent slope.
5. γ_{13} is the effect of the main independent variable on the latent intercept.

Under the assumption of MVN errors in both the measurement and structural model, the distribution of $\mathbf{Y}$ is MVN as well. The mean for any given Y_j is:

$$\begin{aligned} \mathrm{E}(Y_j) = (b_{13}\gamma_{31}) + \gamma_{12} z_1 + (\gamma_{13} + b_{13}\gamma_{33}) z_2 + \\ (b_{23}\gamma_{31}) t_j + (\gamma_{23} + b_{23}\gamma_{33}) z_2 t_j \end{aligned}$$

and the covariance matrix for the vector of Y_j's is:

$$\mathrm{Cov}(\mathbf{Y}) = \mathbf{W \Psi W}^{\mathrm{T}} + \mathbf{\Sigma}$$

where

$$\mathbf{W} = \begin{bmatrix} 1 & t_1 & b_{13} + t_1 b_{23} \\ 1 & t_2 & b_{13} + t_2 b_{23} \\ \vdots & \vdots & \vdots \\ 1 & t_6 & b_{13} + t_6 b_{23} \end{bmatrix}.$$

The primary setting we considered in our simulations assumed a constant effect of heavy alcohol use (the main independent variable of interest) over time. If there is no alcohol by time interaction (i.e. $\gamma_{23} = 0$ and $b_{23} = 0$), the total effect of heavy alcohol use on CD4 count is represented by $\gamma_{13} + b_{13}\gamma_{33}$, which is the sum of its direct (γ_{13}) and indirect ($b_{13}\gamma_{33}$) effect through ART adherence.

Secondary analyses assuming the effect of heavy alcohol use changes over time were also performed. In this setting, the interaction between alcohol and time is the primary interest. The total effect of alcohol use on the change of CD4 count over time is represented by coefficients corresponding to the interaction between alcohol use and time, $\gamma_{23} + b_{23}\gamma_{33}$, the sum of the direct and indirect effects, respectively.

Linear Mixed Effects Models

The LMM can be used to evaluate the total effect of heavy alcohol consumption on CD4 cell count, however, the mediated relationship is not explicitly modeled with

a single LMM. Two mixed models were considered for comparison to the SEM. The first model, which we will refer to as LMM1, includes the mediator (x) as a covariate in the model. In the formula below we have again eliminated the subject index (i) for simplicity and let j = 1,..., 6 index time-point. The first model is:

$$Y_j = \beta_0 + \beta_1 z_1 + \beta_2 z_2 + \beta_3 t_j + \beta_4 x + \beta_5 z_2 t_j + r_1 + r_2 t_j + e_j \quad (6)$$

where β_j's are unknown regression coefficients relating covariates to the mean of Y_j, r_1 and r_2 denote the random intercept and random slope, respectively and e_j is the random error with zero mean representing deviation of responses from the corresponding predicted means. In matrix form $\mathbf{e} = (e_1, e_2,..., e_6)^T$ is the vector of unknown random errors with E($\mathbf{e}$) = $\mathbf{0}$ and Cov($\mathbf{e}$) = $\mathbf{E}$; $\mathbf{r} = (r_1, r_2)^T$ is the vector of the random intercept and slope coefficients with E($\mathbf{r}$) = $\mathbf{0}$ and Cov($\mathbf{r}$) = $\mathbf{G}$. In the primary setting we explore, where the effect of alcohol is assumed constant and therefore the alcohol by time interaction in (6) is excluded, the effect of the main independent variable on the outcome is represented by β_2. In the secondary setting explored where there is an alcohol by time interaction, the parameter of interest is β_5. Variables in the causal pathway are often omitted as they otherwise artificially attenuate the effect of the main independent variable of interest. Thus, we also considered a second mixed model, which we refer to as LMM2, where we refit the model given in Equation 6 excluding the term for ART adherence ($\beta_4 x$).

The LMM assumes MVN errors, therefore the distribution of $\mathbf{Y}$ is MVN as well. For the LMM1 in Equation (6), the mean for any given Y_j is:

$$\mathrm{E}(Y_j) = \beta_0 + \beta_1 z_1 + \beta_2 z_2 + \beta_3 t_j + \beta_4 x + \beta_5 z_2 t_j$$

and the covariance matrix of the vector of Y_j's is:

$$\mathrm{Cov}(\mathbf{Y}) = \mathbf{Z}\,\mathbf{G}\,\mathbf{Z}^{\mathrm{T}} + \mathbf{E}$$

where

$$\mathbf{Z} = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_6 \end{bmatrix}.$$

In a comparison of the means of the SEM and LMM under the assumption of multivariate normality, there are two notable differences. First, the mean for the SEM is explicitly modeled as a function of both the direct and indirect effects of the main independent variable z_2 (i.e. $\gamma_{13} + b_{13}\,\gamma_{33}$) whereas in the LMM it is simply a function of its total effect (i.e. β_2). Second, in a LMM that includes the mediating variable (LMM1), the mean of the outcome is a function of the mediating variable itself, but this is not the case for the SEM where the mean depends instead on the effect of the mediating variable on the latent intercept and slope (i.e., b_{13} and b_{23}). With regard to covariance matrices, in the LMM the covariance depends on the values of time and the covariance matrix of the latent intercept and slope. In contrast, the covariance from the SEM depends explicitly on the mediating variable and its effects. That is, it is a function of the values of time, the parameters associated with the effect of the mediating variable on the random intercept and random slope, as well as the covariance matrix of the latent intercept, slope, and the mediating variable. Thus, the magnitude of the mediated effect and the covariance of the mediator and latent variables influence the covariance of Y in the SEM but this is not the case in the LMMs.

We simulated data under the mediated SEM and then fit the data with both the LMM1 and LMM2 models. The SEM model was also fit as a reference standard to compare with the LMM results. The goal was to identify advantages and disadvantages of using the LMM relative to the SEM in a longitudinal data setting where a mediator was present. In evaluating model performance, we focused on the parameters representing the total effect of heavy alcohol use on CD4 cell count.

Data Characteristics to be Varied in Simulated Data

We simulated datasets assuming the mediated longitudinal relationship described in the path diagram (Figure 2). The data were generated under the SEM model as the objective was to evaluate the performance of the mixed model when the SEM is expected to be preferable. The factors we evaluated were sample size, effect size and distributional assumptions.

Sample Size

A range of sample sizes was evaluated. We considered sample sizes as small as 25 and increased values up to 500 at which point both the SEM and LMM performed well.

Effect Size and Total Effect Distribution

When the effect of heavy alcohol use is constant over time, the total effect of the main independent variable on the latent intercept is given by $\gamma_{13} + b_{13}\,\gamma_{33}$. We defined the effect size by scaling this quantity by the total standard deviation of the latent intercept:

$$\text{Effect Size : Intercept} = \frac{\gamma_{13} + b_{13}\,\gamma_{33}}{\sqrt{\mathrm{var}(\mathrm{U_{1i}})}} \quad (7)$$

where

$$\text{var}(U_{1i}) = \gamma_{12}^2 \text{var}(z_{1i}) + \gamma_{13}^2 \text{var}(z_{2i}) + b_{13}^2 \gamma_{33}^2 \text{var}(z_{2i}) + \text{var}(\zeta_{1i}) + b_{13}^2 \text{var}(\zeta_{3i}) + b_{13} \text{cov}(\zeta_{1i}, \zeta_{3i}).$$

When the effect of heavy alcohol use changes over time, the direct effect of the main independent variable on the latent intercept is set to zero, and thus the effect size on the latent slope was defined as:

$$\text{Effect Size : Slope} = \frac{\gamma_{23} + b_{23}\,\gamma_{33}}{\sqrt{\text{var}(U_{2i})}} \quad (8)$$

where

$$\text{var}(U_{2i}) = \gamma_{22}^2 \text{var}(z_{1i}) + \gamma_{23}^2 \text{var}(z_{2i}) + b_{23}^2 \gamma_{33}^2 \text{var}(z_{2i}) + \text{var}(\zeta_{2i}) + b_{23}^2 \text{var}(\zeta_{3i}) + b_{23} \text{cov}(\zeta_{2i}, \zeta_{3i}).$$

We examined a range of effect sizes including small (approximately 0.2), medium (approximately 0.5) and large (approximately 0.8), as defined by Cohen [19]. In addition, within each effect size we varied the distribution of the direct and indirect effects and explored the following three scenarios: equally distributed direct and indirect effects; primarily direct effect; primarily indirect effect.

Distribution of the Outcome Variable

Both SEMs and LMMs assume normally distributed errors of the outcome variables. We compared the performance of each type of model when this assumption was not met. For each distribution evaluated we considered two scenarios: i) only the errors from the measurement model were non-normal and ii) errors of both the measurement and structural models were non-normal. The following distributions were evaluated:

1. Uniform($-\sqrt{3}, \sqrt{3}$) distribution- the parameters of the uniform distribution were chosen to obtain a mean of zero and a variance of one to be comparable to the standard normal setting.
2. Lognormal(0, $\sqrt{0.4812}$) distribution- the lognormal parameters were chosen such that the mean of the residual errors was equal to zero and the variance was equal to one. To achieve a mean of zero, exp(0.4812/2) was subtracted from all generated lognormal values.
3. Contaminated normal(0.4, 10) distribution- a mixture of a standard normal and normal with variance of 10, where 40% of the data were from the latter distribution [20].
4. Fleishman/Mattson method. The Fleishman [21] method describes a way to generate non-normal random variates with known skewness and kurtosis. The Mattson [22] method provides a way to generate non-normal random variates with specified correlation from non-normal random variates with known skewness and kurtosis. The method also provides a formula for the skewness and kurtosis of the randomly generated correlated values. The combined method [23] allowed us to change only the distribution of the errors while keeping the overall variance and correlation the same. We used two Fleishman/Mattson distributions. The first had a moderate level of skewness and low level of kurtosis. The second was highly skewed and highly kurtotic. The first results in a variance of 1, a skewness of 0.75 and kurtosis of 0 for the residual errors of the measurement model and skewness of (0.53, 0.5 and 0.75) and kurtosis of (-1.5, -1.9 and -3) for the three residual errors of the structural model, respectively. The second Fleishman/Mattson distribution we used results in a variance of 1, skewness of 1.75 and kurtosis of 3.75 of the residual errors of the measurement model and skewness of (1.2, 1.4 and 1.8) and kurtosis of (0.4, 0.5 and 0.8) for the three residual errors of the structural model, respectively.

Data Simulation

To generate a dataset under the mediated SEM data structure, the following steps were taken.

1. Two multivariate normal random variates were generated, one to be the residual error of the latent intercept and one to be the residual error of the latent slope. When evaluating the impact of distributional assumptions, the non-normal distributions defined in the previous section replaced the multivariate normal distribution in this step. The Mattson method [22] was used to keep the covariance between the random intercept and random slope at the same level as was used for the normal simulations.
2. The value of the latent intercept and latent slope were computed according to the structural model given in Equation 4.
3. Independent normal errors were created to be the residual errors for each of the repeated measures of outcome. When evaluating the impact of distributional assumptions, the non-normal distributions defined in the previous section replaced the multivariate normal distribution in this step.
4. The values of the longitudinal outcome variables were computed according to the measurement model given in Equation 3.
5. Steps 1 through 4 were repeated 1000 times to create 1000 datasets.

6. Each generated dataset was fit with the SEM, the LMM1 (i.e. with the mediator as a covariate), and the LMM2 (i.e. without the mediator as a covariate).
7. Model performance was assessed with the following: i.) Bias- the difference between the true parameter value and the mean observed parameter value divided by the true parameter value. ii.) Coverage probability - the percentage of the 1000 95% confidence intervals that contained the true parameter value. iii.) Power - the percentage of the 1000 datasets in which a hypothesis test of the parameter of interest was statistically significant. With a sample size of 1000, and a true power of 80%, the width of a 95% confidence interval around a power estimate based on the simulations would be approximately 5.0 percentage points. For a true 95% coverage probability, the width of a 95% confidence interval around a coverage probability estimate would be approximately 2.7 percentage points.

Results

The results from the mixed effects models focus primarily on the models that do not adjust for the mediator (i.e., LMM2) because these models capture the total effect of the main independent variable. Results from the mixed model adjusting for the mediator (LMM1) appear to capture the direct rather than total effect of the primary independent variable and are therefore only included in the sample size results to demonstrate this result. However, since the primary objective of the analysis was to evaluate the total effect of the main independent variable, we present only the comparison of the SEM and the mixed model that excludes the mediator in the remainder of the results. Typically, variables associated with the outcome are included in a model, including independent predictors and confounders. However, because mediators are in the causal pathway, it is recommended that such variables be excluded from a model to avoid attenuating the true association between an exposure and outcome [24]. Thus the LMM2 model is consistent with the general practice of excluding as a covariate variables thought to be in the causal pathway.

Sample Size

The results from the sample size variation are displayed in Table 1. With sample sizes of 25, 50 and 100, the estimated power to detect the total effect for all models was quite low (14%-65%). We note that for the SEM, with a sample size of 100, the power for the total effect was 65%, while the power to detect the direct and indirect effects were 26% and 71% (data not shown), respectively. With a sample size of 200, the power for the SEM and LMM2 were both high and similar in magnitude (91% and 89%, respectively), although the power for the LMM1 remained low at 41%. The estimated power of the LMM1 for a sample size of 500 was less than 80%. With a sample size of 500 the power for the SEM and LMM2 models were 99%, therefore we did not evaluate larger sample sizes.

The estimated coverage probabilities for the total effect for all sample sizes for the SEM were high, ranging from 92% to 96%. For the LMM2, these coverage probabilities range from 91% to 95%. For the LMM1, the coverage probabilities were much lower than either the SEM or the LMM2 and decreased with increasing sample size. This trend is likely due to the large bias of the estimate for the LMM1 and resulting confidence intervals that are not centered at the true value. Thus the wider confidence intervals from smaller sample sizes are more likely to include the true value.

Because the model performance was good for the SEM and LMM2 with a sample size of 250 subsequent simulations evaluating effect sizes and distributional assumptions were conducted using this sample size.

Table 1 Impact of sample size.

Simulated Data	Mediated SEM			LMM with Mediator as Covariate LMM1			LMM without Mediator LMM2		
Sample Size	Bias (%)	Coverage Probability (%)	Power (%)	Bias (%)	Coverage Probability (%)	Power (%)	Bias (%)	Coverage Probability (%)	Power (%)
25	1.5	92	26	-48	85	14	2.1	91	26
50	1.5	93	41	-49	84	16	2.0	93	42
100	1.2	94	65	-49	75	26	1.7	94	65
200	-1.2	96	91	-51	56	41	-1.1	95	89
250	-1.6	94	95	-52	48	47	-1.6	93	94
500	-1.0	95	100	-51	21	78	-1.0	95	100

Impact of sample size on model performance in evaluating the total effect of the main independent variable on random intercept.
Based on 1000 simulated datasets with medium effect size equally distributed between direct and indirect effects.

Effect Size and Total Effect Distribution

The results from the set of simulation studies varying the effect of the main independent variable on the random intercept are displayed in Table 2.

In evaluating the effect of the main independent variable on the random intercept, the power for the SEM and the LMM2 (i.e. LMM without the mediating variable) was > 99% when the effect size was large, regardless of whether the effect was primarily direct, primarily indirect or equally distributed between the direct and indirect paths. The coverage probabilities were also very similar between the SEM and LMM2 (≥ 93% in all cases).

For a medium effect, the point estimate of power was slightly higher for the SEM compared to the LMM2 regardless of how the effect was distributed. For example, when the effect was equally distributed the power was 94% for the LMM2 and 95% for the SEM. Although the power for all models was high (≥ 92%), the power for the SEM and the LMM2 appeared to increase as the proportion of the direct effect increased. When the effect was primarily indirect the power was 92% for the LMM2 and 93% for the SEM. The coverage probabilities for the total effect were again ≥ 93% for both the SEM and the LMM2. The higher point estimates of power in the SEM appeared to be due to the larger standard error of the effect estimate in the LMM2. Similar trends were observed with the medium-small and small effect sizes although the power for all models dropped markedly with the small effect size. For example, power was approximately 32% for both models in the case of a small effect size, equally distributed between direct and indirect effects.

Distributions

In simulations evaluating the effect of distributional assumptions, we used a sample size of 250 and a medium-small effect size that was equally distributed in direct and indirect effects (see Table 3). Results from the model with a normal distribution (and the same sample size and effect distribution as described above) had power of 80% and bias of -2.0% for both the SEM and the LMM2 and a coverage probability of 94% for the SEM and 93% for the LMM2. This is referred to below as the normal comparison model.

Assuming a uniform distribution on the residual errors of the measurement model the power to detect the total effect was estimated to be 82% for the SEM and 81% for the LMM2. Both of these models had similar estimates of power which were slightly greater than the power estimate of the comparison models with normal residual errors. This was likely due to an underestimation of the standard error of the parameters. For the model with normal residual errors, the mean of the standard errors of the total effect was 0.16 whereas the mean of the standard errors for the uniform was 0.15. The coverage probability was very similar for both the SEM and the LMM2 (96%). These estimates were slightly higher than those for the normal comparison models. The bias was small (< 1%) for both SEM and LMM2. The results were similar when a uniform distribution was used for the residual errors of both the measurement and structural models.

Model performance was good overall for both the SEM and LMM2 when errors followed a log-normal distribution (see Table 3), although not as good as under the uniform distribution.

Table 2 Impact of effect size and effect distribution.

Simulated Data			Mediated SEM			Mixed Model without Mediator		
Effect Size	Effect Distribution	Effect	Bias (%)	Coverage Probability (%)	Power (%)	Bias (%)	Coverage Probability (%)	Power (%)
Large	Equal	Total	-0.8	94	100	-0.8	94	100
	Direct	Total	-0.8	94	100	-0.8	93	100
	Indirect	Total	0.9	94	100	-0.6	94	100
Medium	Equal	Total	-1.5	94	95	-1.5	93	94
	Direct	Total	-1.5	94	97	-1.5	93	96
	Indirect	Total	-1.3	94	93	-1.3	93	92
Medium-Small	Equal	Total	-2.0	94	80	-2.0	93	80
	Direct	Total	-0.9	94	84	-0.9	93	83
	Indirect	Total	-1.8	94	80	-1.8	93	79
Small	Equal	Total	-3.7	94	32	-3.7	93	32
	Direct	Total	-3.9	94	36	-4.0	93	35
	Indirect	Total	-3.7	94	34	-3.7	93	33

The impact of effect size and its distribution on model performance in evaluating the total effect of the main independent variable on the random intercept. Based on 1000 simulated datasets with a sample size of 250.

Table 3 Impact of distributional assumptions.

Simulated Data		Mediated SEM			Mixed Model without Mediator		
Distribution	Non-normal Residual Error	Bias (%)	Coverage Probability (%)	Power (%)	Bias (%)	Coverage Probability (%)	Power (%)
Uniform	Measurement	0.7	96	82	-0.8	96	81
	Measurement & Structural	1.4	97	87	1.6	96	85
Log-normal	Measurement	-1.3	94	82	-1.3	94	80
	Measurement & Structural	-0.7	95	82	-0.7	95	81
Contaminated Normal	Measurement	-0.2	95	18	0.07	95	17
	Measurement & Structural	9.5	96	8	11.1	95	8
Fleishman/Mattson 1	Measurement & Structural	-2.9	94	79	-3.0	94	78
Fleishman/Mattson 2	Measurement & Structural	-2.4	94	80	-2.6	94	78

The impact of distributional assumptions on model performance in evaluating the total effect of the main independent variable on the random intercept.
Based on 1000 simulated datasets with a medium-small effect size, equally distributed and a sample size of 250.
The Fleishman/Mattson 1 distribution is moderately skewed and slightly kurtotic.
The Fleishman/Mattson 2 distribution is highly skewed and kurtotic.

Power declined noticeably for both models when errors of the measurement model followed a contaminated normal distribution, however, the coverage probability and bias remained good (power ≤ 18%, coverage probability ≥ 95% and bias ≤ 0.2%). The coverage probability remained high likely due to the large standard error estimates yielding wide 95% confidence intervals. Similar trends were observed for the contaminated normal distribution on the residual errors of the measurement and structural models although model performance declined for both the SEM and LMM2. The lower power of both the SEM and the LMM2 fit to the contaminated normal data may be explained by the relatively large values of the residual variances created by the contaminated normal. For example, the estimated mean values for the residual variance was 40.6 in the SEM and the LMM2 compared to around 1 in the models based on a normal distribution. The effect of a large residual variances is a decrease in the true effect size of the main independent variable.

Both models performed well when measurement and structural errors followed Fleishman/Mattson distribution. The power was similar for the SEM and LMM2 (79% and 78%, respectively). Both models had the same coverage probabilities and bias, 94% and -3%, respectively. Similar values and trends were seen with the second Fleishman/Mattson distribution.

The results of an SEM are generally presented with at least two fit indices [25]. Commonly used fit statistics are the chi-square statistic, the AIC, the root mean square error of approximation (RMSEA) and the standardized root mean square residual (SRMR). LMM models are not usually presented with fit statistics, although during model specification, fit indices like the log-likelihood, AIC and BIC have been used for model selection [26]. The fit statistics for the SEM and LMM2 under different error distributions for both the measurement and structural models are given in Table 4.

Lower values suggest better model fit for each of the fit statistics presented. RMSEA values of < 0.05 and SRMR values of < 0.1 are considered good fit [25]. However, all of the fit statistics from the SEM with the exception of the AIC indicate that the log-normal, the most skewed distribution, had the worst fit followed by the second Fleishman/Mattson model which also has a skewed distribution. The AIC from the SEM and all of the fit statistics from the LMM2 were less affected by the skewness of the log-normal.

Results when Effects of the Main Independent Variable Change Over Time

The simulation results for the effect of the main independent variable on the random slope did not differ qualitatively from the results for the random intercept (data not shown). Overall, the point estimates of power were slightly higher for all models (SEM and LMM) likely due to the fact that no covariate by time interaction was included in the models. The lack of this additional interaction term results in higher true power for the effect of the main independent variable on the random slope. In general, the simulation results were similar to those observed in the primary setting where the main independent variable had a constant effect across time.

Discussion

Linear mixed effects models are often used to analyze longitudinal data. Although LMMs can be applied in settings where mediation is present, it is unclear whether they perform sufficiently well relative to SEMs which have a framework that explicitly allows for

Table 4 Goodness of fit.

		Distribution			
Model	Fit Statistic	Uniform	Lognormal	Fleishman/Mattson 1	Fleishman/Mattson 2
SEM	Chi-square	36.6	58.0	38.4	43.7
	RMSEA	0.010	0.040	0.023	0.021
	SRMR	0.039	0.043	0.041	0.041
	AIC	7309	7288	7307	7307
Mixed Model without Mediator	-2LogLikelihood	5550	5537	5549	5547
	AIC	5566	5553	5565	5563
	BIC	5594	5581	5593	5591

Assessing goodness of fit of SEMs and LMMs with datasets with non-normal error distributions.

Normal comparison model has SEM values of Chi-square = 38.2, RMSEA = 0.012, SRMR = 0.040, AIC = 7306 and LMM values of Negative 2Loglikelihood = 5548, AIC = 5564 and BIC = 5592.

mediational analyses. The objective of this paper was to evaluate the performance of the LMM in the analysis of longitudinal data with a single mediating variable, a setting conducive to the use of SEMs.

The simulation studies were conducted to assess whether the mixed effects model adequately modeled the mediated longitudinal relationships or if employing SEMs was necessary. The LMM and SEM were compared under a range of settings evaluating sample size, effect size and distributional assumptions. The results of our simulation study suggest that the mixed effects model performs comparably to the SEM with respect to power, bias and coverage probability in the analysis when the objective is to estimate the total effect of a primary independent variable. In addition, we demonstrated that mixed effects models used for the purpose of estimating total effects should not include mediating variables as covariates, since resulting coefficients represent the direct effect of the main independent variable on the outcome and erroneous conclusions could be drawn if these effects were interpreted as the total effect.

Both the LMM and SEM were robust to violations of the normality assumption. For the SEM, lack of normality had a larger impact on the model fit statistics than on power, coverage probability and bias. The uniform distribution, an example of a kurtotic, but not skewed distributions, had little effect on the SEM fit statistics compared to the negative impact observed for the lognormal and Fleishman/Mattson distributions, which were both skewed and kurtotic. Generally, the highest levels of skewness had the worst fit. For this reason, caution should be used in applying the SEM when the normality of the data is in question, particularly if the distribution of the data is skewed.

There are several considerations in deciding whether to use an SEM or a mixed model to analyze longitudinal data when a mediating factor is present. The SEM may provide a marginal increase in power, although the difference may not be statistically significant. More importantly, it efficiently evaluates the mechanism of the total effect, decomposing direct and indirect pathways, in a single model. However, larger sample sizes are required to make inferences about the specific direct and indirect effects. If the sample size is limited and the goal is to evaluate only the total effect of a primary independent variable, rather than delineating direct versus indirect effects, then the mixed model provides similar power and coverage probability to the SEM. Although direct and indirect effects could be evaluated by fitting additional mixed models (e.g. models with and without mediating factors), it is a less efficient approach compared to the SEM. In addition, there are broader issues that may influence choice of model such as clinical context, study design and sample size.

Complex SEMs may be difficult to implement without specialized software. Although common software packages such as SAS and R have the capability to run SEMs, software designed specifically for SEMs (e.g. Mplus, LISREL and AMOS) may be more intuitive and user-friendly in model specification, particularly in the development of highly complex models.

The current study examines one specific setting of mediated longitudinal data. Other situations with different data structures where mediation is present could also be explored, e.g. situations where the mediator and the primary independent variable as well as the outcome are repeatedly measured, categorical outcomes, and settings with more complex pathways between variables. In addition, we specifically explored the question of whether the LMM performs sufficiently in a setting favorable to the SEM. Future studies examining broader settings where the data arise from non-SEMs would provide further insight into the use of the LMM and SEM in mediated longitudinal settings.

Conclusions

In general, both SEMs and LMMs were robust methods with similar power in a variety of scenarios. The main

advantage of the SEM is the ability to estimate the direct and indirect pathways of the effect of the primary independent variable on the outcome, given sufficient sample sizes. Despite not directly modeling the mediated pathways, LMMs excluding mediating variables performed well with respect to power, bias and coverage probability in modeling the total effect of the primary independent variable on the outcome.

Author details
[1]Department of Biostatistics, Boston University School of Public Health, 715 Albany Street Crosstown Center-3rd Floor Boston, MA 02118, USA. [2]Clinical Research Program, Children's Hospital Boston and Harvard Medical School, 300 Longwood Avenue Boston, MA 02115, USA.

Authors' contributions
All authors read and approved the final manuscript. EAB was involved in the conception of the study, designing and performing the simulation analysis and drafting the manuscript. HC and TH were involved in conception of the study, designing the simulation study and critically revising the manuscript. DMC was involved in conception of the study, designing the simulation analysis and drafting the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Received: 16 October 2009 Accepted: 19 February 2010
Published: 19 February 2010

References

1. Laird N, Ware J: **Random effects models for longitudinal data.** *Biometrics* 1982, **38**:963-974.
2. Krull JL, MacKinnon DP: **Multilevel Modeling of Individual and Group Level Mediated Effects.** *Multivariate Behavioral Research* 2001, **36**:249-277.
3. Baron RM, Kenny DA: **The moderator-mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations.** *Journal of Personality and Social Psychology* 1986, **51**:1173-1182.
4. McArdle JJ, Epstein D: **Latent growth curves within developmental structural equation models.** *Child Development* 1987, **58**:110-133.
5. Muthen B: *Best Methods for the Analysis of Change: Recent Advances, Unanswered Questions, Future Directions* Washington, D.C.: APA 1991 chap. Analysis of Longitudinal Data Using Latent Variable Models with Varying Parameters1-17.
6. Muthen BO, Curran PJ: **General Longitudinal Modeling of Individual Differences in Experimental Designs: A Latent Variable Framework for Analysis and Power Estimation.** *Psychological Methods* 1997, **2(4)**:371-402.
7. Rovine MJ, Molenaar PCM: **A Structural Modeling Approach to a Multilevel Random Coefficients Model.** *Multivariate Behavioral Research* 2000, **35**:51-88.
8. Rovine MJ, Molenaar PCM: *New methods for the analysis of change* Washington, D.C.: APA 2001 chap. A Structural Equations Modeling Approach to the General Linear Mixed Model67-96.
9. McArdle JJ, Aber MS: *Statistical Methods in Longitudinal Research* New York: Academic Press 1990 chap. Patterns of change within latent variable structural equations models151-224.
10. Mehta PD, West SG: **Putting the Individual Back Into Individual Growth Curves.** *Psychological Methods* 2000, **5**:23-43.
11. Curran PJ: **Have Multilevel Models Been Structural Equation Models All Along?.** *Multivariate Behavioral Research* 2003, **38(4)**:529-569.
12. Bauer DJ: **Estimating Multilevel Linear Models as Structural Equation Models.** *Journal of Educational and Behavioral Statistics* 2003, **28(2)**:135-167.
13. Stoel RD, Wittenboer van Den G, Hox J: **Estimating Multilevel Linear Models as Structural Equation Models.** *Metodologia de las Ciencias del Comportamiento* 2003, **5**:21-42.
14. Mehta PD, Neale MC: **People Are Variables Too: Multilevel Structural Equations Modeling.** *Psychological Methods* 2005, **10(3)**:259-284.
15. Sanchez BN, Budtz-Jorgensen E, Ryan LM, Hu H: **Structural Equation Models: A Review with Applications to Environmental Epidemiology.** *Journal of the American Statistical Association* 2005, **100(472)**:1443-1455.
16. Bollen KA: *Structural Equations With Latent Variables* New York, New York: Wiley 1989.
17. Samet J, Cheng D, Libman H, Nunes D, Alperen J, Saitz R: **Alcohol consumption and HIV disease progression.** *Journal of Acquired Immune Deficiency Syndrome* 2007, **46(2)**:194-199.
18. Meredith W, Tisak J: **Latent Curve Analysis.** *Psychometrika* 1990, **55**:107-122.
19. Cohen J: *Statistical Power Analysis for the Behavioral Sciences* Hillsdale, New Jersey: Lawrence Erlbaum Associates 1988.
20. Tukey J: *Contributions to Probability and Statistics* Palo Alto, California: Stanford University Press 1960 chap. A Survey of Sampling from Contaminated Distributions448-485.
21. Fleishman AI: **A Method for Simulating Non-normal Distributions.** *Psychometrika* 1978, **43(4)**:521-532.
22. Mattson S: **How to Generate Non-normal Data for Simulation of Structural Equation Models.** *Multivariate Behavioral Research* 1997, **32(4)**:355-373.
23. Reinartz WJ, Echambadi R, Chin WW: **Generative Non-normal Data for Simulation of Structural Equation Models Using Mattson's Method.** *Multivariate Behavioral Research* 2002, **37(2)**:227-244.
24. Rosner B: *Fundamentals of Biostatistics* Pacific Grove, California: Dubxbury Press 2005.
25. Kline RB: *Principles and Practice of Structural Equation Modeling* New York, New York: Guilford Press, second 2005.
26. Fitzmaurice GM, Laird NM, Ware JH: *Applied Longitudinal Analysis* Hoboken, New Jersey: John Wiley and Sons, Inc 2004.

Pre-publication history
The pre-publication history for this paper can be accessed here:http://www.biomedcentral.com/1471-2288/10/16/prepub

doi:10.1186/1471-2288-10-16
Cite this article as: Blood *et al.*: Performance of mixed effects models in the analysis of mediated longitudinal data. *BMC Medical Research Methodology* 2010 **10**:16.

Comparison of In-Hospital Versus 30-Day Mortality Assessments for Selected Medical Conditions

Ann M. Borzecki, MD, MPH,†‡ Cindy L. Christiansen, PhD,*† Priscilla Chew, MPH,* Susan Loveland, MAT,†§ and Amy K. Rosen, PhD†§¶*

Background: In-hospital mortality measures such as the Agency for Healthcare Research and Quality (AHRQ) Inpatient Quality Indicators (IQIs) are easily derived using hospital discharge abstracts and publicly available software. However, hospital assessments based on a 30-day postadmission interval might be more accurate given potential differences in facility discharge practices.

Objectives: To compare in-hospital and 30-day mortality rates for 6 medical conditions using the AHRQ IQI software.

Methods: We used IQI software (v3.1) and 2004–2007 Veterans Health Administration (VA) discharge and Vital Status files to derive 4-year facility-level in-hospital and 30-day observed mortality rates and observed/expected ratios (O/Es) for admissions with a principal diagnosis of acute myocardial infarction, congestive heart failure, stroke, gastrointestinal hemorrhage, hip fracture, and pneumonia. We standardized software-calculated O/Es to the VA population and compared O/Es and outlier status across sites using correlation, observed agreement, and kappas.

Results: Of 119 facilities, in-hospital versus 30-day mortality O/E correlations were generally high (median: $r = 0.78$; range: 0.31–0.86). Examining outlier status, observed agreement was high (median: 84.7%, 80.7%–89.1%). Kappas showed at least moderate agreement ($k > 0.40$) for all indicators except stroke and hip fracture ($k \leq 0.22$). Across indicators, few sites changed from a high to nonoutlier or low outlier, or vice versa (median: 10, range: 7–13).

Conclusions: The AHRQ IQI software can be easily adapted to generate 30-day mortality rates. Although 30-day mortality has better face validity as a hospital performance measure than in-hospital mortality, site assessments were similar despite the definition used. Thus, the measure selected for internal benchmarking should primarily depend on the healthcare system's data linkage capabilities.

From the *Center for Health Quality, Outcomes, and Economic Research, Bedford VAMC, Bedford, MA; †Department of Health Policy and Management, Boston University School of Public Health, Boston, MA; ‡Section of General Internal Medicine, Boston University School of Medicine, Boston, MA; §Center for Organization, Leadership, and Management Research, VA Boston Healthcare System, Boston MA; and ¶Department of Family Medicine, Boston University School of Medicine, Boston, MA.

Supported by Veterans Health Administration.

Reprints: Ann M. Borzecki, MD, MPH, Center for Health Quality, Outcomes, and Economic Research, Bedford VA Hospital (152), 200 Springs Rd, Bedford, MA 01730. E-mail: amb@bu.edu.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.lww-medicalcare.com).

ISSN: 0025-7079/10/4812-1117

Key Words: hospital mortality, quality indicators, quality improvement

(*Med Care* 2010;48: 1117–1121)

Both in-hospital and 30-day mortality rates from procedures or medical conditions are frequently used for assessing quality of hospital care. Although in-hospital measures are relatively straightforward to derive, requiring only hospital discharge abstracts, critics of such measures cite potential biases due to differences in facilities' lengths of stay and discharge practices.[1,2] Commonly used measures for quality improvement (QI) and public reporting include the Agency for Healthcare Research and Quality (AHRQ) mortality indicators, a subset of the Inpatient Quality Indicators (IQIs), and the Centers for Medicare and Medicaid (CMS) 30-day mortality measures for acute myocardial infarction (AMI), congestive heart failure (CHF), and pneumonia.[3–5]

The IQIs were developed in response to demand for reliable, easy to use quality measures that could be applied across healthcare systems and settings. IQI mortality rates for several procedures and conditions can be readily obtained using publicly available software.[6] However, they were developed on the Health Care Utilization Project (HCUP) dataset which lacks linkage to information occurring outside the index admission. Although use of 30-day mortality measures allows for standardization of follow-up time, additional data sources and linkage capabilities are required that many hospital/healthcare systems may lack. Thus, using the IQIs for local benchmarking and QI remains an attractive option. This option would be even more attractive if performance assessments using in-hospital mortality were comparable to those obtained using 30-day mortality.

The Veterans Health Administration (VA) is the nation's largest integrated healthcare system, offering care to almost 7 million veterans. Unlike much of the private sector, the VA has the ability to link multiple datasets which are used for administrative and research purposes. This capability represents a unique opportunity to apply the AHRQ IQIs to VA data to compare in-hospital and 30-day medical condition mortality rates, and to determine whether facility-based assessments of standardized mortality rates vary by method.

METHODS

Study Population

Our study population consisted of veterans discharged from VA inpatient care during fiscal year (FY) 2004 through 2007 with a principal diagnosis of AMI, CHF, stroke, gastrointestinal (GI) hemorrhage, hip fracture, and pneumonia.

Data Sources

Our main data source was the National Patient Care Database's Patient Treatment File (PTF) which contains information on all VA discharges.[7] It includes demographics, diagnoses (principal and secondary ICD-9-CM codes), procedures (ICD-9-CM codes), and discharge status. Additional vital status information was obtained from the VA's Vital Status files.[8]

Overview of the Inpatient Quality Indicators

IQI development is described in detail elsewhere.[3] The final IQI set resulted from a 4-step process including comprehensive literature review, structured clinical panel review, coding expert consultation, and empirical analyses of potential IQIs. It includes both procedure- and medical condition-related mortality indicators which are in-hospital mortality rates associated with specific procedures or conditions that have shown provider variation and where evidence exists that high mortality may be associated with poorer care.[3] Originally intended as screens and case-finding tools for local QI efforts, these indicators are being used increasingly for hospital profiling and public reporting. The National Quality Forum has endorsed several of the mortality IQIs as hospital performance measures and CMS are adding selected mortality IQIs to their hospital reporting initiative.[9,10]

Given we previously found relatively little facility variation for the procedure-related mortality IQIs, herein we focus on the medical condition-related IQIs.[11]

Analyses

Analyses were performed using AHRQ IQI software (v.3.1), All Patient Refined-Diagnostic Related Group (APR-DRG) software (3M, v.24; for risk adjustment), both downloadable from AHRQ's website, and the Statistical Analysis System (SAS, v.8.0).[6,12] Our unit of analysis was the individual hospitalization.

We applied the IQI and APR-DRG software to the PTF (FY04–07) to generate in-hospital observed and risk-adjusted mortality rates (ie, deaths per 100 discharges with the specified principal diagnosis) and ratios of observed to expected rates (O/Es) at the level specified (eg, VA-wide or hospital/facility-level). The IQI software calculates a patient's expected probability of death during the hospital admission using parameter estimates derived from logistic regression models run on the HCUP population that include age, sex, age-sex interactions and APR-DRG mortality risk score as covariates. These regression estimates are used as true parameter values for the admission-level covariates from the population of interest. (patient-level observations are assumed to be independent). The sum of the expected probability of death across all qualifying hospital admissions is the number of deaths expected (E) if care were similar to that in the HCUP system. The ratio of the observed number of deaths (O) to the expected number of deaths (E) is a standardized mortality ratio, indirectly standardized to the HCUP population (Appendix A, online only, available at: http://links.lww.com/MLR/A129, for further details on IQI models.).

We used IQI software-generated O/Es and calculated 95% confidence intervals (CIs). Because the IQIs use an expected rate based on the HCUP population, we standardized facility-level O/Es (and CIs) to the overall VA rate in the 4-year period by multiplying by a constant equal to the inverse of the VA's national O/E. Sites were considered outliers if the 95% CI did not include 1.0. We similarly determined 30-day mortality standardized O/Es after linking the PTF and Vital Status files. For a given indicator, subjects with more than 1 IQI-related admission within a 30-day period and who died within 30-days of the original admission were only counted once in the 30-day numerator.

We compared in-hospital and 30-day median observed rates using Wilcoxon rank sum tests and standardized mortality O/Es using correlation coefficients. We calculated agreement with respect to outlier status via observed agreement (concordance) and weighted kappas.[13] Facility-level O/E pairs were considered concordant if there was no difference in facility assessment by mortality method versus discordant if there was a change.

RESULTS

Table 1 shows facility-level sample characteristics. Overall, the mean age was 69.5 (s.d.: 12.2); the sample was predominantly male (98%) and white (64%). Comorbidities were relatively prevalent; 61% of discharges had hypertension; over a third had coronary artery disease, diabetes, or chronic lung disease.

All medical conditions had significantly higher observed 30-day mortality rates compared with in-hospital (Table 1). Correlations between in-hospital and 30-day mortality O/Es showed strong positive associations (ie, coefficients ≥ 0.70; $P < 0.05$) except for hip fracture ($r = 0.31$, $P < 0.05$). Measures of agreement using weighted kappa based on outlier status followed similar trends as correlations, being at least moderate ($k > 0.40$) for all IQIs except hip fracture and stroke which showed slight ($k = 0.12$) and fair agreement ($k = 0.22$), respectively (Fig. 1).

Simple observed agreement or concordance between paired data did not necessarily follow the same trends as kappas. Median observed agreement was high at 0.88, ranging from 0.81 for pneumonia to 0.89 for GI hemorrhage and hip fracture, despite it having the lowest kappa (Table 2). The median number of facilities across indicators that changed outlier status was 18 (range: 12–23); the number of sites changing status was highest for pneumonia.

Examining across indicators by discordant pairs, facilities were slightly more likely to change from a nonoutlier based on in-hospital mortality to a low or high outlier using 30-day mortality. A change in outlier status from high to low or vice versa was very rare. This occurred for only 1 facility

TABLE 1. Facility-Level Study Sample Characteristics*

Facility Level Variable	AMI† N = 30,893	CHF N = 88,874	Stroke N = 26,065	GI Hem N = 39,063	Hip Fracture N = 8855	PNA N = 75,049
Admissions, n, median (25th, 75th percentile)	209 (54, 395)	652 (321, 1027)	173 (78, 310)	290 (149, 429)	72 (23, 118)	536 (356, 874)
In-hospital deaths, %, median (25th, 75th percentile)	7.2 (5.1, 11.8)	3.4 (2.7, 4.5)	6.0 (4.2, 7.7)	2.6 (1.6, 3.6)	5.1 (1.8, 7.8)	5.7 (4.5, 7.9)
30-d deaths, %, median (25th, 75th percentile)	11.1 (8.2, 16.1)	6.8 (5.9, 8.3)	9.0 (6.7, 11.5)	4.9 (3.7, 5.9)	9.5 (5.8, 12.9)	9.5 (8.3, 11.1)
Age, median (25th, 75th percentile)	69.5 (66.0, 74.5)	72.0 (70.0, 74.0)	69.0 (66.0, 73.0)	70.0 (66.0, 72.0)	78.0 (76.0, 80.0)	71.0 (69.0, 73.0)
Male sex, %, median (25th, 75th percentile)	98.7 (97.8, 99.4)	98.5 (97.9, 99.1)	97.9 (97.0, 98.8)	97.9 (97.0, 98.6)	96.4 (94.3, 98.4)	97.6 (96.9, 98.0)
Race, %, median (25th, 75th percentile)						
White	70.7 (56.7, 83.4)	67.2 (52.5, 80.7)	61.9 (48.8, 75.0)	66.4 (49.4, 78.3)	72.2 (60.3, 83.9)	72.1 (57.5, 82.0)
Black	11.5 (4.1, 24.1)	14.4 (4.9, 34.1)	20.0 (7.7, 36.8)	13.3 (5.1, 30.5)	10.4 (4.0, 20.8)	9.5 (3.5, 23.6)
Hispanic	1.8 (0.6, 6.9)	0.6 (0.2, 3.1)	2.5 (0.6, 5.9)	1.5 (0.4, 6.1)	5.1 (2.4, 7.6)	0.6 (0.2, 2.8)
Other	11.5 (6.1, 22.2)	6.8 (3.0, 17.7)	14.2 (6.1, 26.5)	9.6 (4.8, 21.9)	14.3 (6.8, 28.6)	8.8 (4.0, 22.1)
Selected comorbidities, %, median (25th, 75th percentile)‡						
CVD	8.4 (6.5, 10.8)	6.6 (5.2, 8.3)	—	5.9 (4.4, 7.6)	10.4 (6.8, 14.3)	6.2 (4.8, 7.7)
Chronic lung disease	24.9 (20.0, 31.0)	40.3 (35.0, 47.0)	16.2 (12.0, 20.6)	21.8 (18.8, 27.0)	26.1 (20.0, 33.3)	43.5 (49.5, 56.1)
CHF	34.1 (27.1, 40.9)	—	11.6 (8.2, 14.0)	14.7 (12.5, 18.3)	13.4 (8.3, 18.0)	21.6 (17.6, 24.8)
CAD	—	58.7 (52.7, 65.6)	27.5 (22.4, 32.2)	28.0 (22.2, 33.2)	23.7 (17.1, 30.4)	28.1 (21.7, 32.7)
Depression	6.0 (3.8, 8.1)	8.1 (6.0, 10.2)	6.5 (4.2, 8.9)	6.5 (4.5, 9.1)	5.6 (2.3, 8.1)	8.9 (6.3, 11.7)
Diabetes mellitus	41.7 (37.5, 45.1)	51.9 (47.0, 55.9)	37.6 (34.0, 40.8)	30.5 (26.3, 33.2)	24.8 (20.0, 28.2)	29.5 (26.7, 32.5)
Hypertension	60.2 (53.6, 66.2)	59.8 (50.4, 67.9)	71.4 (65.6, 75.3)	51.8 (45.5, 58.0)	48.5 (42.2, 56.0)	49.7 (45.1, 53.7)

All observed 30-d death rates are significantly higher than in-hospital rates ($P < 0.05$).

*Analysis done at hospitalization/admission level. Medical conditions are defined per AHRQ IQI indicator specifications.[14] (See Appendix B, online only, available at http://links.lww.com/MLR/A130).

†There are 2 AMI mortality IQIs. We found similar results for AMI mortality with and AMI without transfers, we therefore present data for AMI with transfers only.[14]

‡Comorbidities are based on secondary diagnoses from the index admission using the HCUP Comorbidity Software; codes for CAD and CVD are based on prior work.[15,16]

AMI indicates acute myocardial infarction; CHF, congestive heart failure; GI Hem, gastrointestinal hemorrhage; PNA, pneumonia; CAD, coronary artery disease; CVD, cerebrovascular disease.

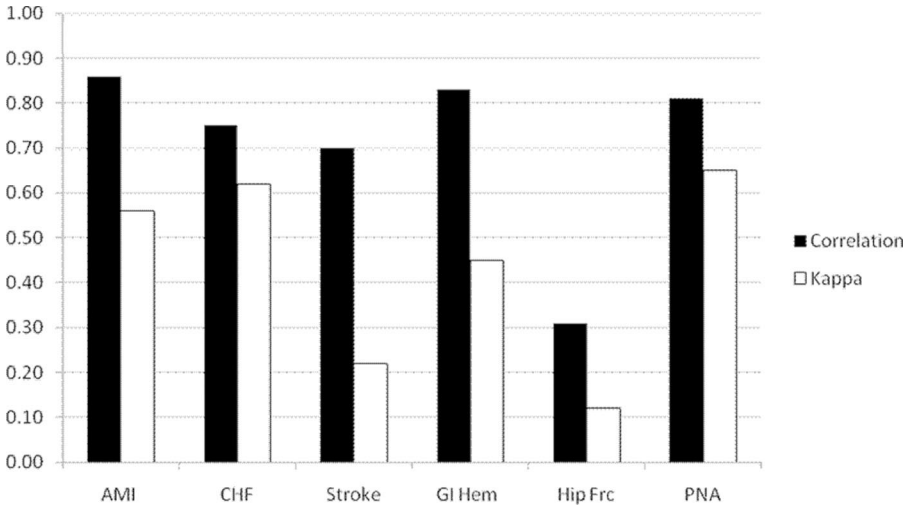


FIGURE 1. Facility-level correlations and kappas. Correlation coefficients for in-hospital versus 30-day standardized O/Es are shown. All correlations were significant ($P < 0.05$). Kappa measures of agreement are based on assessment of outlier status as low outlier, not an outlier or high outlier. Kappa interpretation: <0 = poor, 0 to 0.20 = slight, 0.21 to 0.40 = fair, 0.41 to 0.60 = moderate, 0.61 to 0.80 = substantial, 0.81 to 1.00 = almost perfect agreement. AMI indicates acute myocardial infarction; CHF, congestive heart failure; GI hem, gastrointestinal hemorrhage.

and 1 indicator, pneumonia; the site changed from a high to a low outlier (Table 2.) Facilities were more likely to change from a low or nonoutlier to a high outlier for 4 indicators (CHF, stroke, GI hemorrhage, hip fracture; Table 2). The median number of facilities across indicators that changed status from high to a nonoutlier or low outlier, or vice versa was 10 (range, 7–13); the number of sites changing status was highest for stroke.

DISCUSSION

This is the first study to use the IQIs to compare in-hospital and 30-day mortality across several medical con-

TABLE 2. In-Hospital Versus 30-d Mortality—Change in Outlier Status

Outlier Status (Site Pairs), n (%)	AMI N = 116	CHF N = 119	Stroke N = 119	GI Hem N = 119	Hip Fracture N = 110	PNA N = 119
Concordant pairs*	**102 (87.9)**	**97 (81.5)**	**97 (81.5)**	**106 (89.1)**	**98 (89.1)**	**96 (80.7)**
Not an outlier	92 (79.3)	74 (62.2)	93 (78.2)	100 (84.0)	97 (88.2)	67 (56.3)
Low-low	5 (4.3)	12 (10.1)	1 (0.8)	4 (3.4)	1 (0.9)	13 (10.9)
High-high	5 (4.3)	11 (9.2)	3 (2.5)	2 (1.7)	0 (0)	16 (13.4)
Discordant pairs	**14 (12.1)**	**22 (18.5)**	**22 (18.5)**	**12 (10.9)**	**12 (10.9)**	**23 (19.3)**
Low vs. not	2 (1.7)	3 (2.5)	7 (5.9)	1 (0.8)	1 (0.9)	9 (7.6)
Not vs. low	2 (1.7)	8 (6.7)	2 (1.7)	4 (3.4)	4 (3.6)	4 (3.4)
High vs. not	5 (4.3)	5 (4.2)	4 (3.4)	3 (2.5)	3 (2.7)	4 (3.4)
Not vs. high	5 (4.3)	6 (5.0)	9 (7.6)	5 (4.2)	4 (3.6)	5 (4.2)
High vs. low	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (0.8)
Low vs. high	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)

Concordant pairs are sites that have no change in their outlier status; discordant pairs do have a change in their outlier status based on the assessment method. Direction of change for discordant pairs is indicated starting from in-hospital to 30-d, eg, for "Low versus Not," for AMI, 2 sites changed from low outlier status based on in-hospital mortality O/Es to not being considered as outliers based on 30-d mortality O/Es.

*The percentage of concordant pairs is the same as observed agreement.

AMI indicates acute myocardial infarction; CHF, congestive heart failure; GI Hem, gastrointestinal hemorrhage; PNA, pneumonia.

ditions. Using a large national sample of VA admissions, we found that the AHRQ IQI software can be readily adapted to generate 30-day mortality rates. However, assessments of hospital outlier status comparing standardized in-hospital and 30-day mortality rates were similar regardless of the indicator. Whereas correlation coefficients and outlier status agreement based on kappas were lowest for stroke and hip fracture, based on observed agreement, the worst agreement was seen for pneumonia and best for hip fracture and GI hemorrhage. Although, at most 19% of facilities changed status on any indicator, with facilities more likely to change from a non-outlier using in-hospital rates to an outlier using 30-day rates. Potential mislabeling of sites as high outliers when they were low outliers or average, or as average or low outliers when they were high, was relatively uncommon, occurring in approximately 10% of facilities for any given indicator.

This is also the first study comparing mortality measurement methods to use relatively recent data that better reflect current discharge practices, and to use a VA population. The consistency of our findings with older studies in non-VA populations suggests our results are not dependent on the specific methodology used. Rosenthal et al compared in-hospital and 30-day standardized mortality ratios in Medicare discharges with CHF in 30 Ohio hospitals during 1992 to 1994.[17] Their correlation between mortality ratios was similar to ours at 0.78. Seven hospitals changed outlier status (ie, 77% observed agreement). On the basis of their findings, they concluded that in-hospital mortality is a "reasonably valid marker for 30-day mortality." Chassin et al also found generally high hospital-level correlations between in-hospital and 30-day mortality for 48 surgical and medical conditions (0.54 for all conditions), using 1984 Medicare data.[18] However, AMI and CHF correlations were highest (0.79 and 0.71, respectively).

As noted, the IQI software uses HCUP-based expected rates which likely differ from the VA population, given typical VA and non-VA population differences (eg, more males and comorbidities in veterans).[19] Other healthcare systems that differ from the HCUP population will face similar case-mix issues. Standardizing software-derived O/Es to the population of interest, as we have done, enables meaningful within system determination of outlier status among sites. Further, the IQI software uses parameter estimates from logistic models that do not account for patient clustering within hospitals. However, other investigators have found nearly identical results with respect to facility-level surgical morbidity and mortality outlier status when comparing hierarchical and logistic regression models.[20]

Although 30-day measures are generally considered more accurate measures of hospital performance than in-hospital mortality measures because they are less dependent on hospital discharge practices, the reviewed studies and our findings suggest that this is less of a concern except perhaps for stroke and hip fracture which had lower kappas and correlation coefficients. These conditions are both fairly dependent on inpatient rehabilitation services. Therefore, as found in non-VA settings, we presume that lengths of stay may be highly variable across facilities, depending on access to in-house versus out-of-house VA and non-VA rehabilitation facilities.[21–23] Thus, some patients may end up dying in-hospital because of relatively long stays, whereas others may die in a non-VA subacute care facility. This raises some concern over use of the hip fracture IQI by CMS for hospital reporting.[10]

Further, 30-day mortality measures like those used by CMS are also considered more accurate than in-hospital measures like the IQIs, because of more robust risk adjustment (they use diagnoses from the 12 months preceding the admission). CMS currently generates hospital-level assessments with respect to these measures using hospital-submitted Medicare claims and reports them on their Hospital Compare website.[24] Whereas CMS' methods may be replicated, software to generate these measures is not readily available and only 3 conditions are represented (AMI, CHF, and pneumonia).[4,5] Although the IQIs use the APR-DRGs and secondary diagnoses from the index admission for comor-

bidity risk adjustment, previous studies using alternative risk-adjustment methods based either on administrative data alone or more robust risk-adjustment with clinical data found similar results with respect to hospital mortality assessments.[17,18]

Thus, the AHRQ IQI software represents a useful option for calculating 30-day mortality across various medical conditions, especially for systems that have linkage to out-of-hospital death data. Although 30-day mortality is considered to have better face validity as a measure of hospital performance, for systems wanting to do internal benchmarking that do not have ready access to out-of-hospital death data, our findings suggest that similar results for selected conditions can be obtained by using in-hospital data alone.

REFERENCES

1. Baker DW, Einstadter D, Thomas CL, et al. Mortality trends during a program that publicly reported hospital performance. *Med Care*. 2002; 40:879–890.
2. Kahn KL, Brook RH, Draper D, et al. Interpreting hospital mortality data. How can we proceed? *JAMA*. 1988;260:3625–3628.
3. Agency for Healthcare Research and Quality. *Guide to Inpatient Quality Indicators.* June 2002. Version 3.0 (February 20, 2006). Rockville, MD: Department of Health and Human Services, Agency for Healthcare Research and Quality; 2006.
4. Krumholz HM, Normand SL, Bratzler DW, et al. Risk-adjustment methodology for hospital/surveillance and public reporting: supplement no. 1: 30-day mortality model for pneumonia.Oklahoma City, OK: Oklahoma Foundation for medical Quality; 2006.
5. Krumholz HM, Normand SL, Galusha DH, et al. Risk-adjustment models for AMI and HF 30-day mortality: methodology. Woodlawn, MD: Centers for Medicare and Medicaid Services; 2005.
6. AHRQ Quality Indicators Software version 3.1. Available at: http://www.qualityindicators. ahrq.gov/software.htm. Accessed February 1, 2009.
7. 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.
8. Arnold N, Sohn M, Maynard C, et al. VIReC Technical Report 2: VA NDI Mortality Data Merge Project. Hines, IL: VA Information Resource Center; 2006.
9. National Quality Forum. NQF Endorsed Standards. Available at: http://www.qualityforum.org/Measures_List.aspx. Accessed January 10, 2010.
10. CMS. 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. Borzecki AM, Chew P, Loveland S, et al. Inpatient quality indicators: comparison of rates in a large integrated healthcare system. Washington, DC: Academy Health Meeting; 2008.
12. AHRQ Quality Indicators 3M APR-DRG V24 Limited License Grouper SAS Documentation. Available at: http://www.qualityindicators.ahrq.gov/software.htm. Accessed January 10, 2010.
13. Sackett DL, Haynes RB, Guyatt GH, et al. *Clinical Epidemiology: A Basic Science For Clinical Medicine*. 2nd ed. Boston, MA: Little Brown and Company; 1991.
14. Agency for Healthcare Research and Quality. *Inpatient Quality Indicators: Technical Specifications*. June 2002. Version 3.1 (March 12, 2007). Rockville, MD: Department of Health and Human Services, Agency for Healthcare Research and Quality; 2007.
15. HCUP Comorbidity Software. Available at: http://www.hcup-us.ahrq.gov/toolssoftware/comorbidity/comorbidity.jsp#overview. Accessed January 10, 2010.
16. 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.
17. Rosenthal GE, Baker DW, Norris DG, et al. Relationships between in-hospital and 30-day standardized hospital mortality: implications for profiling hospitals. *Health Serv Res*. 2000;34:1449–1468.
18. Chassin MR, Park RE, Lohr KN, et al. Differences among hospitals in Medicare patient mortality. *Health Serv Res*. 1989;24:1–31.
19. Agha Z, Lofgren RP, VanRuiswyk JV, et al. Are patients at Veterans Affairs medical centers sicker? A comparative analysis of health status and medical resource use. *Arch Intern Med*. 2000;160:3252–3257.
20. Cohen ME, Dimick JB, Bilimoria KY, et al. Risk adjustment in the American College of Surgeons National Surgical Quality Improvement Program: a comparison of logistic versus hierarchical modeling. *J Am Coll Surg*. 2009;209:687–693.
21. Beech R, Ratcliffe M, Tilling K, et al. Hospital services for stroke care. A European Perspective. European Study of Stroke Care. *Stroke*. 1996; 27:1958–1964.
22. Parker MJ, Todd CJ, Palmer CR, et al. Inter-hospital variations in length of hospital stay following hip fracture. *Age Ageing*. 1998;27:333–337.
23. Jencks SF, Williams DK, Kay TL. Assessing hospital-associated deaths from discharge data. The role of length of stay and comorbidities. *JAMA*. 1988;260:2240–2246.
24. US Department of Health and Human Services. Hospital Compare—A quality tool provided by Medicare. Available at: http://www.hospitalcompare.hhs.gov/Hospital/Search/Welcome.asp?version=default&browser=IE%7C6%7CWinXP&language=English&defaultstatus=0&pagelist=Home. Accessed January 10, 2010.

Trends in the Inpatient Quality Indicators

The Veterans Health Administration Experience

Ann M. Borzecki, MD, MPH,†‡ Cindy L. Christiansen, PhD,*† Susan Loveland, MAT,†§ Priscilla Chew, MPH,* and Amy K. Rosen, PhD†§¶*

Background: The Agency for Healthcare Research and Quality Inpatient Quality Indicators (IQIs), which include in-hospital mortality and utilization rates, have received little attention in the Veterans Health Administration (VA), despite extensive private sector use for quality improvement.

Objectives: We examined the following: the feasibility of applying the IQIs to VA data; temporal trends in national VA IQI rates; temporal and regional IQI trends in geographic areas defined by Veterans Integrated Service Networks' (VISNs); and VA versus non-VA (Nationwide Inpatient Sample) temporal trends.

Methods: We derived VA- and VISN-level IQI observed rates, risk-adjusted rates, and observed to expected ratios (O/Es), using VA inpatient data (2004–2007). We examined the trends in VA- and VISN-level rates using weighted linear regression, variation in VISN-level O/Es, and compared VA to non-VA trends.

Results: VA in-hospital mortality rates from selected medical conditions (stroke, hip fracture, pneumonia) decreased significantly over time; procedure-related mortality rates were unchanged. Laparoscopic cholecystectomy rates increased significantly. A few VISNs were consistently high or low outliers for the medical-related mortality IQIs. Within any given year, utilization indicators, especially cardiac catheterization and cholecystectomy, showed the most inter-VISN variation. Compared with the non-VA, VA medical-related mortality rates for the above-mentioned conditions decreased more rapidly, whereas laparascopic cholecystectomy rates rose more steeply.

Conclusions: The IQIs are easily applied to VA administrative data. They can be useful to tracks rate trends over time, reveal variation between sites, and for trend comparisons with other healthcare systems. By identifying potential quality events related to mortality and utilization, they may complement existing VA quality improvement initiatives.

Key Words: quality of care, administrative data, quality improvement

(*Med Care* 2010;48: 694–702)

The Agency for Healthcare Research and Quality (AHRQ) Inpatient Quality Indicators (IQIs) were developed in response to demand for reliable, valid, easy to use quality measures that could be used across healthcare systems and settings.[1] They are evidence-based indicators derived from hospital discharge data. They include mortality, utilization, and volume indicators designed to screen for potential inpatient quality problems, highlighting areas in which quality of care should be further investigated.[2]

Originally intended for use as screens and case-finding tools for local quality improvement (QI) efforts, the IQIs are increasingly being used as performance measures for hospital profiling and public reporting.[3–7] Given their reliance on administrative data, AHRQ offers caution about such use.[3] Notwithstanding this, the National Quality Forum recently endorsed several of these indicators as hospital performance measures, and the Centers for Medicare and Medicaid Services are adding select individual and 2 composite mortality IQIs to their hospital reporting initiative.[7–9]

The IQIs were developed on Healthcare Cost and Utilization Project (HCUP) data, representing private sector hospitals from a subset of states. Thus, their validity and utility in systems, such as the Veterans Health Administration (VA), which may differ from HCUP with respect to patient case-mix or coding practices, is unknown. As the nation's largest integrated healthcare system the VA is a national leader in QI innovations and provision of high quality care. This has occurred through means such as development of a highly integrated, comprehensive electronic medical record, performance monitoring of certain chronic and acute conditions, programs aimed at improving surgical care, and more recently, the Quality Enhancement Research Initiative, intended to facilitate research translation into practice for selected patient populations.[10–12] Many of these initiatives, such as the National Surgical Quality Improvement Program (NSQIP), rely on chart abstraction and are thus labor-inten-

From the *Center for Health Quality, Outcomes and Economic Research, Bedford VAMC, Bedford, MA; †Department of Health Policy and Management, Boston University School of Public Health, Boston, MA; ‡Section of General Internal Medicine, Boston University School of Medicine, Boston, MA; §VA Boston Healthcare System, West Roxbury, MA; and ¶Department of Family Medicine, Boston University School of Medicine, Boston, MA.

Supported by Veterans Health Administration.

Reprints: Ann M. Borzecki, MD, MPH, CHQOER, Bedford VA Hospital (152), 200 Springs Rd, Bedford, MA 01730. E-mail: amb@bu.edu.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.lww-medicalcare.com).

ISSN: 0025-7079/10/4808-0694

sive and expensive.[11] The IQIs, which have thus far received little VA attention, represent a potentially cost-efficient tool to complement these efforts.

Further, in the private sector there is evidence of decreasing rates of several of the mortality-related IQIs over time,[13] as well as geographic variation in IQI rates.[6,14] Whether similar trends exist in the VA is unknown. Although prior VA studies have found variation in longer-term survival for specific IQI-represented conditions, such as acute myocardial infarction (AMI), plus other measures of care,[15,16] the degree of variation across the VA has not been studied using the IQIs. Such variation, if found, will help improve care by focusing VA QI efforts on specific conditions or indicators.

This study's purpose was to examine: (1) the feasibility of applying the IQIs to VA data, (2) temporal trends in national VA IQI rates, (3) temporal trends and variation in rates across VA geographic regions, represented by Veterans Integrated Service Networks (VISNs), and (4) differences in VA versus non-VA temporal trends. Lessons learned from our experience may be applicable to other healthcare systems.

METHODS

Study Population

Our study population consisted of veterans discharged from VA inpatient care during fiscal year (FY) 2004 through 2007.

Data Collection

Our main data source was the National Patient Care Database's Patient Treatment File which contains information on all VA discharges.[17] It includes demographics, diagnoses (principal and secondary ICD-9-CM codes), surgical and nonsurgical procedures (ICD-9-CM coded), and discharge status. We obtained supplemental comorbidity information from the National Patient Care Database outpatient file[18] and facility information from the VA Support Service Center Occupancy Rate Reports (bed counts) and VA Office of Academic Affiliations (resident counts). National Inpatient Sample (NIS) IQI rates and standard errors were obtained from the HCUP-net site.[14]

Overview of the Inpatient Quality Indicators

IQI development followed a four-step process including comprehensive literature review, structured clinical panel review, coding expert consultation, and empirical analyses of potential IQIs.[19] They include 3 types of indicators: (1) Mortality indicators are in-hospital mortality rates associated with specific conditions or procedures that have shown provider variation and where evidence exists that high mortality may be associated with poorer care.[2,15,19–25] (2) Utilization indicators are rates of procedures where concerns exist about overuse, underuse, or misuse.[2,19,26–28] (3) Volume indicators are counts representing proxy quality measures based on evidence that hospitals performing more, highly complex procedures tend to have better outcomes, (eg, survival) for those procedures.[2,19,20,22] To provide more meaningful comparisons, we focused on indicators representing rates, rather than simple volume counts. We omitted 4 delivery-related utilization indicators because of low volume. (see Appendix 1, online only, Supplemental Digital Content 1, available at: http://links.lww.com/MLR/A86, for definitions of the 17 IQIs used in this study).

Applying the IQIs to VA Data: Required Data Elements

The IQIs algorithms link diagnoses and procedure codes with other information contained in standardized hospital discharge data to generate counts and rates.[2] Required data elements include age, sex, admission source, patient disposition, Diagnosis-Related Group (DRG), ICD-9-CM principal and secondary diagnosis codes, ICD-9-CM procedure codes, and All Patient Refined Diagnostic Related Group (APR-DRG) category, severity and mortality scores.[29] Because VA databases differ from HCUP's with respect to structure and specific data elements, we recoded certain variables based on previous work with AHRQ's Patient Safety Indicators.[30] For example, admission source has 19 possible VA values but only 5 in HCUP.

Analyses

Analyses were performed using AHRQ IQI software (v.3.1), APR-DRG software (3M, v.24), and the Statistical Analysis System (SAS, v.8.0).[29,31] Our unit of analysis was the individual hospitalization.

We applied the IQI and APR-DRG software to the Patient Treatment File to generate observed, risk-adjusted rates, and ratios of observed to expected rates (O/Es) for all IQIs at the level specified (ie, VA-wide, and VISN-level). The IQI software uses parameter estimates derived from logistic regression models (for mortality indicators) and linear regression models (for utilization indicators) run on the HCUP population that include age, sex, age-sex interactions, APR-DRG mortality risk score, and APR-DRG severity score as covariates. These regression estimates are used as true parameter values for the admission-level covariates from the population of interest. Patient-level observations are assumed to be independent). Risk-adjusted rates reflect the estimated performance on each IQI for a provider assuming that provider had the "average" case mix among all hospitals in the HCUP estimation sample.[29] Per AHRQ's recommendations, we excluded VISNs with fewer than 3 cases in the denominator for a given year.[29] Notably, higher rates for the utilization indicators potentially indicate poorer quality except for laparoscopic cholecystectomy.

We first examined FY04 through FY07 trends in national VA IQI risk-adjusted rates (software-produced), using weighted linear regression models and calculating *P*-values for their slope. The VA is geographically divided into 21 regional healthcare units known as VISNs. We chose this analysis level rather than the hospital level because many of the surgical procedures included in the mortality IQIs are performed at fewer than half of VA hospitals, whereas all are fully represented at the VISN level except for esophageal resection and pancreatic resection.

We examined VISN-level temporal trends in IQI risk-adjusted rates using similar methods to our VA-level models. We compared variation across VISNs over time (FY04–07)

and space, using software-generated O/Es and calculated 95% confidence intervals (CIs). Because the IQIs are risk-adjusted using an expected rate based on the HCUP population, we standardized the O/Es (and CIs) to the overall VA rate in each year by multiplying by a constant equal to the inverse of the VA's national O/E. VISNs were identified as outliers if the 95% CI did not include 1.0. The IQI software also generates composite scores for medical and procedure-related mortality which are output as risk ratios.[32] We generated composite scores using the default weighting of denominator weights, calculated 95% CIs, standardized scores to the VA population by considering the mean VA score as 1.0, and examined for outliers.

We next compared VA to non-VA trends using HCUP's posted NIS data.[14] Because we had access to FY03 VA data, and HCUP data were available for calendar years 2003 through 2007, to improve trend comparisons, we added an additional year of data to this analysis. Because of HCUP and VA data differences (eg, HCUP weights results to the total number of NIS discharges, VA data are not discharge-weighted) and differences in sample sizes and precision of estimates, we ran separate weighted linear regression models of the risk-adjusted IQI rates (one for NIS and one for VA), with weights calculated as 1/standard error.[2] We then compared slope estimates (FY03-FY07) using *t*-tests.

RESULTS

Table 1 shows sample characteristics. The total number of hospitalizations (FY04-FY07) was over 2 million; 18% of these met IQI definitions. Comparing all VA discharges to IQI-related discharges, the mean age was 63.0 (±13.6) years versus 68.3 (±11.8); both samples were predominantly male (96%) and white (>60%). IQI-related discharges had more diabetes (39% vs. 34%) and less severe mental illness (11% versus 17%) than the entire discharge population. Of the 123 facilities represented among the 21 VISNs, 83% were teaching hospitals.[33] The median facility bed-size was 126. Appendix 2 (online only, Supplemental Digital Content 2, available at: http://links.lww.com/MLR/A87) shows selected facility characteristics aggregated by VISN.

TABLE 1. Study Sample Characteristics*

	2004–2007 All Discharges N = 2,272,894		2004–2007 IQI-Related Discharges N = 403,828	
	n	%	n	%
Male sex	2,180,853	96.0	395,276	96.0
Race				
White	1,363,607	60.0	252,585	62.6
Black	420,910	18.5	62,580	15.5
Hispanic	34,183	1.5	5606	1.4
Other	454,194	20.0	83,057	20.6
Age group				
18–39	88,748	3.9	2789	0.7
40–64	1,211,298	53.3	160,273	39.7
65–74	426,152	18.8	100,502	24.9
75+	546,696	24.1	140,264	34.7
Selected comorbidities†				
Diabetes mellitus	761,490	33.5	157,500	39.0
HIV	27,643	1.2	3427	0.8
Spinal cord injured	47,981	2.1	3514	0.9
Stroke	143,456	6.3	43,193	10.7
Severely mentally ill‡	385,222	17.0	42,551	10.5

*Analysis done at hospitalization level.
†The NPCD outpatient file was used to identify these selected comorbidities.
‡No. discharges of patients with a preceding diagnosis of schizophrenia, bipolar disorder or depression.

At the national level risk-adjusted rates of all the medical condition-related mortality IQIs decreased over time, although this was statistically significant only for acute stroke, hip fracture and pneumonia (from 10.5 to 8.8, 3.8 to 2.6, and 8.1 to 5.5 deaths per 100 relevant discharges, respectively; Fig. 1A). There were no significant temporal changes in procedure-related mortality IQI rates, despite a slight drop in coronary artery bypass grafting and percutaneous transluminal coronary angioplasty rates (Fig. 1B). Of the utilization indicators, laparascopic cholecystectomy rates increased significantly (from 6.3 to 7.0 per 100 cholecystectomies); the other indicators showed no changes (data not shown; available from authors).

At the VISN-level, most VISNs experienced temporal decreases in medical condition risk-adjusted mortality rates, although this trend was significant for at most 3 VISNs per indicator. No VISN rates increased significantly for any of the medical indicators; (data not shown, available from authors). Similar to overall VA findings, there were no significant trends for the procedure-related mortality IQI rates. For utilization indicators, we only found significant trends for cholecystectomy; whereas most VISNs increased over time, this was significant for just 3 sites.

Of the medical condition-related mortality indicators, pneumonia demonstrated the most inter-VISN variation. Two VISNs (b and n) were consistently high outliers over time (ie, lower 95% CI of the O/E was >1.0 in at least 3 of 4 years), whereas 4 VISNs (l, r, s, u) were consistently low outliers (ie, upper 95% CI was <1.0 in at least 3 of 4 years; Fig. 2A). Figure 3 shows outlier status aggregated across all medical conditions. Four VISNs were consistently high (b, f, i, and n) and 4 (l, r, s, and u) were consistently low outliers on one or more indicators (Fig. 3). By standardized composite scores, 2 VISNs were high (i and n), whereas one was a low outlier (l) in 3 or more years. Using US Census-Bureau area designations, high outlier VISNs tended to be from the South; low outliers were from the West and Midwest.[34]

Examining individual procedure-related mortality indicators, we found at most 2 VISN outliers per indicator per year; these varied by year. There were no outliers on standardized procedure-related composite scores in any years.

For the utilization indicators, cardiac catheterization showed the most variation (Fig. 2B), followed by laparoscopic cholecystectomy. All but 2 VISNs (d and o) were outliers in at least one study year. We found the highest bilateral cardiac catheterization rates in northeastern VISNs

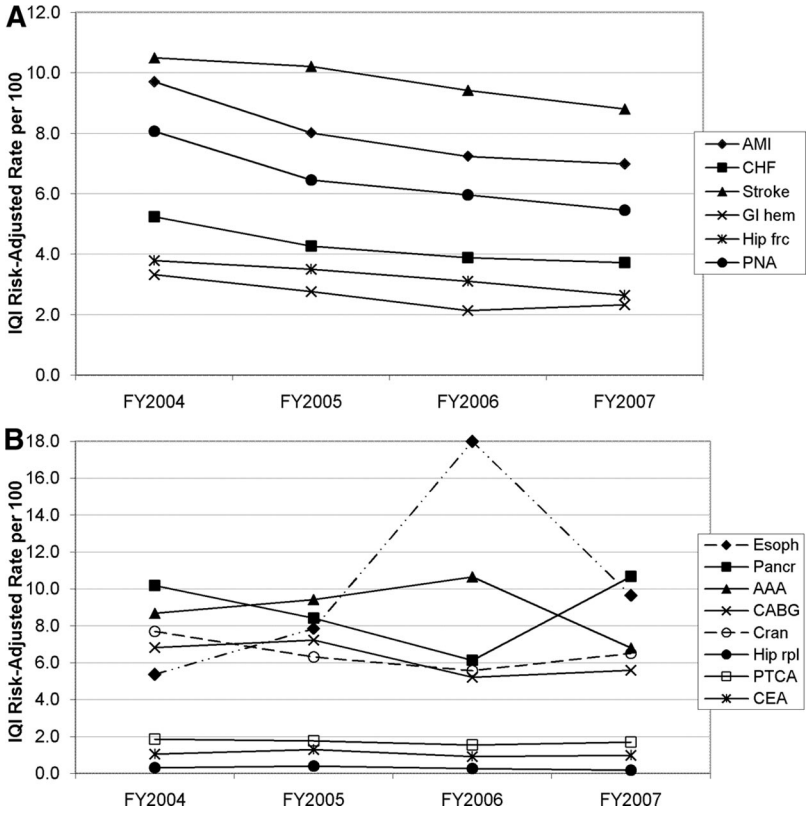


FIGURE 1. National VA Level IQI Mortality Trends. A, Medical Conditions; B, Procedures. Risk-adjusted rates per 100 discharges. Only stroke, hip fracture and pneumonia mortality rates decreased significantly over time; $P < 0.05$. AMI indicates acute myocardial infarction; CHF, congestive heart failure; GI hem, gastrointestinal hemorrhage; Hip frac, hip fracture; PNA, pneumonia; Esoph, esophageal cancer resection; Pancr, pancreatic cancer resection; AAA, abdominal aortic aneurysm repair; CABG, coronary artery bypass graft; Cran, craniotomy; Hip rpl, hip replacement; PTCA, percutaneous transluminal coronary angioplasty; CEA, carotid endarterectomy.

and highest laparascopic cholecystectomy rates in western VISNs.

Comparing VA and NIS risk-adjusted rates from 2003 through 2007, slope estimates differed significantly for AMI, acute stroke, hip fracture and pneumonia mortality, hip replacement, and all 3 utilization indicators (Table 2). AMI, stroke, hip fracture, pneumonia and hip replacement mortality rates, and incidental appendectomy utilization rates declined more rapidly in the VA. Laparascopic cholecystectomy rates rose more steeply, whereas bilateral catheterization rates decreased more slowly, in the VA compared with NIS.

DISCUSSION

This is the first study to apply the AHRQ IQI algorithms to VA data, provide information on VA IQI rates and geographic variation within and across years, and compare VA and non-VA trends. We found these algorithms are easily applied to VA administrative data with minor coding modifications. Our results also suggest that in the VA, the medical condition-related mortality and utilization indicators are more useful than the procedure-related mortality indicators for detecting temporal changes and discriminating between sites.

At the national level, mortality rates for several of the medical condition-related IQIs, (ie, acute stroke, hip fracture, and pneumonia) decreased significantly over time. Of the procedure-related indicators, including mortality and utilization indicators, only laparascopic cholecystectomy changed significantly with utilization rates increasing from FY04 through 07. We found similar trends with respect to improved medical condition survival in the non-VA (NIS) setting, but these changes were significantly greater in the VA. Comparing VISNs, individual utilization indicators showed the most variation, followed by the medical condition-related mortality IQIs; a few VISNs were consistently high or low outliers across years. The procedure-related mortality IQIs demonstrated little variation with respect to temporal trends or outlier status.

Our findings of improved survival from various medical conditions over time are consistent with published literature. An AHRQ-published report using NIS data found improved in-hospital survival between 1994 and 2004 for AMI, stroke, pneumonia, and hip fracture.[13] The authors also noted survival improvement from GI hemorrhage and CHF, plus some of the procedure-related IQIs, including hip replacement (although they sequentially compared rates from one year to the next, instead of using slope estimates). Earlier studies have suggested that in-hospital survival improvements from conditions such as pneumonia may be due to premature discharge and occur at the expense of worsened 30-day or longer term survival.[35,36] However, the Centers for Medicare and Medicaid Services' Medicare Quality Monitoring System reports also showed improved 30-day survival trends for AMI, stroke, and pneumonia (and CHF), between 1992 and 2001.[37]

Two recent studies examined 30-day survival trends following hospitalizations for medical conditions and surgical procedures. This was in the context of studying the impact of resident duty hour reform. They found similar trends of

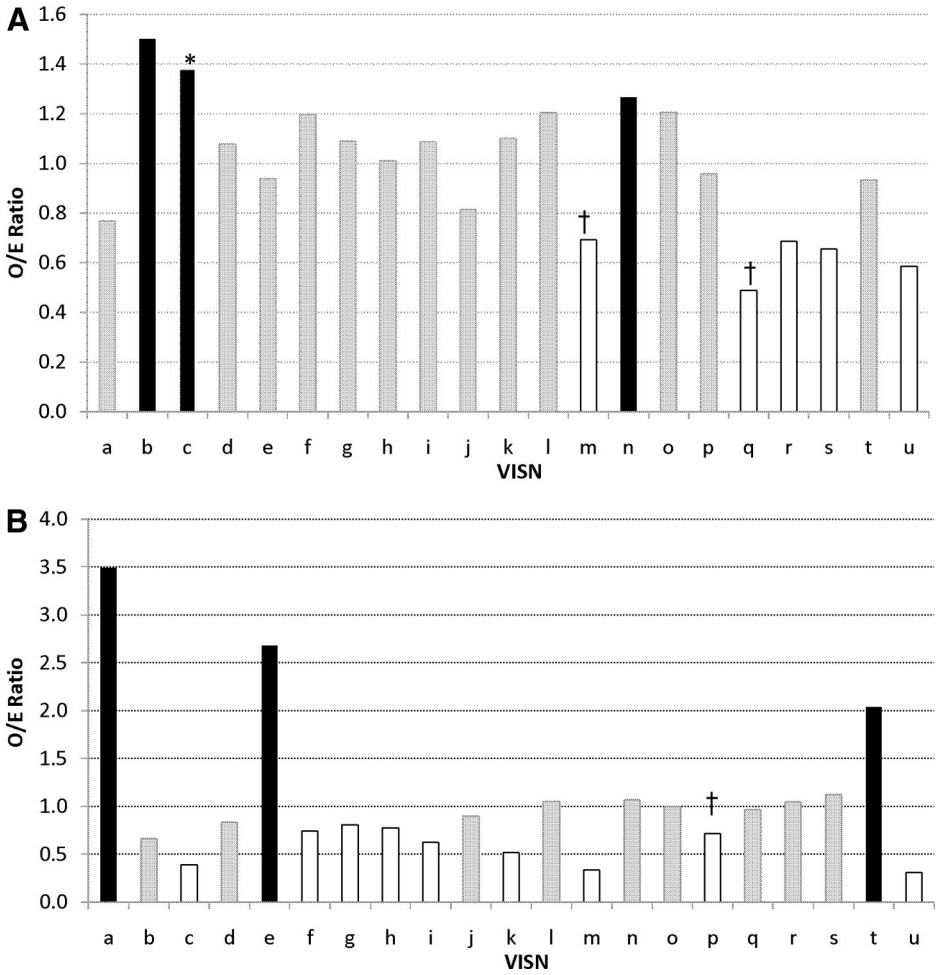


FIGURE 2. FY07 VISN-Level O/E Ratios with Outlier Status. A, Pneumonia; B, Bilateral Cardiac Catheterization. High outliers are indicated by black bars; low outliers by white bars. *High outlier in FY07, but not a consistent outlier. †Low outlier in FY07, but not a consistent outlier. VISN "l" was a low outlier in FY04–06 for pneumonia mortality. Only VISNs "d" and "o" were not outliers for bilateral cardiac catheterization utilization in any study years.

greater survival improvement for medical conditions versus surgical procedures, as well as greater improvement in the VA versus the non-VA (ie, Medicare) population, from July 2000 through June 2005.[38,39] The authors suggest the VA's greater mortality improvement might be because of more teaching hospitals in the VA compared with their non-VA sample; therefore, resident duty reform would be more likely to positively impact such outcomes.[39] Although the continued medical condition survival seen in our study through 2007 may be due to increasing medical service compliance with duty hour rules, this may also be owing to improvements in care resulting from ongoing national VA QI initiatives. Previous studies have found that other VA measures of quality substantially improved and exceeded non-VA settings after VA restructuring in the mid-90s with development of a comprehensive electronic medical record and increased accountability of local managers for performance on quality measures. Jha et al found the VA exceeded Medicare's performance during 1997 through 2000 on several Healthcare Effectiveness Data and Information Set measures that are tracked by the VA's External Peer Review Program (EPRP) including those pertaining to processes of care for patients admitted with AMIs.[10] Another study using a comprehensive assessment containing 294 processes of care, found better quality of care in the VA compared with private managed care patients during the same period.[40]

Our geographic trends are also relatively consistent with existing literature. For the medical mortality IQIs, we found lower outlier VISNs were from the West and Midwest, whereas higher outliers tended to be from the South. Posted HCUP data for calendar years 04 through 06 similarly shows lower death rates in the Mid-West or West; however, the Northeast had the highest death rates for the medical IQIs, not the South.[14] This discrepancy may reflect study population differences. Several geographic variation studies of Medicare patients (which may be more similar to VA patients) have findings similar to ours, with the highest mortality rates in the South for both stroke and AMI.[41,42] A study using 1990s VA data also found the South had the highest AMI mortality.[15] Our utilization findings of highest bilateral cardiac catheterization rates in the Northeast, and highest laparascopic cholecystectomy rates in the West are similar to posted HCUP data.[14]

Designed to be easy to apply, the IQI software uses parameter estimates from logistic and linear regression models based on the HCUP population that do not account for patient clustering within hospitals. Although this is a theoretical methodologic limitation, programs such as NSQIP

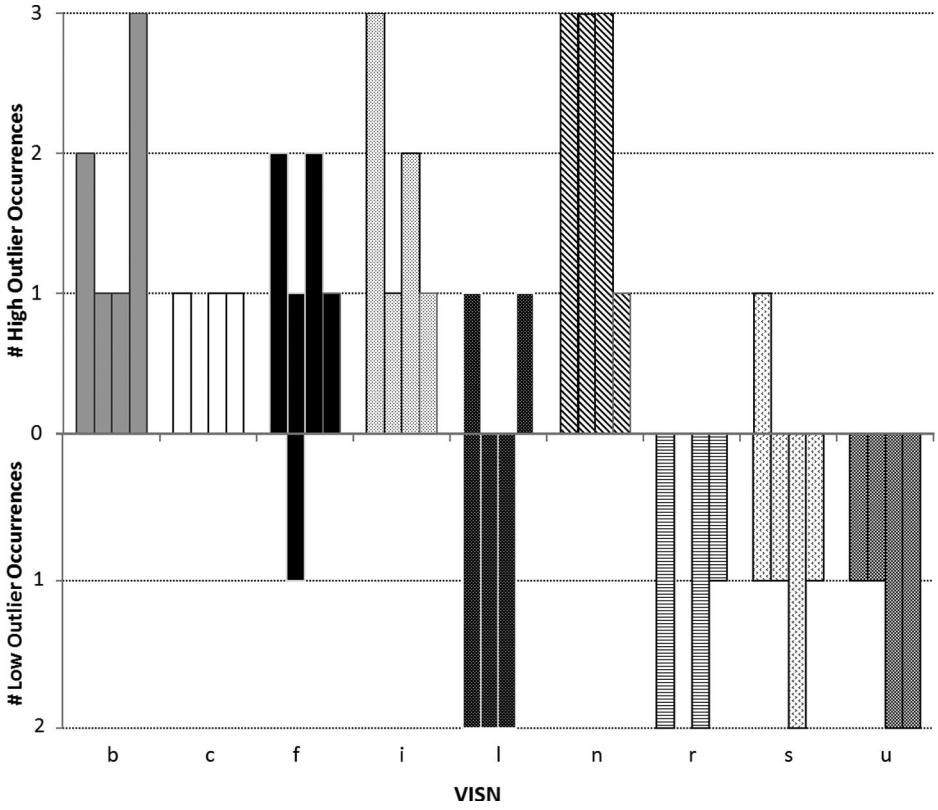


FIGURE 3. Inter-VISN Variation in Medical Condition Related Mortality Indicators. Figure depicts the number of occurrences per year and across years (FY04–FY07) for selected VISNs. Each bar represents one year. Each horizontal gridline represents one occurrence. For example, VISN "l" was a high outlier on one IQI in FY04 and FY07 and a low outlier on 2 IQIs in FY04 through FY06.

similarly use logistic models to generate facility-level surgical outcome O/Es. Hierarchical and logistic regression models using NSQIP data produced nearly identical results.[43]

HCUP-derived expected rates used to generate the risk-adjusted rates are likely to differ from the VA population, given typical VA and non-VA population differences, (eg, more males and more comorbidities in veterans).[44] Therefore, we standardized VISN-level O/Es to VA data to allow more meaningful site comparisons, and were able to identify consistent outlier sites. Other healthcare systems that differ from the HCUP population will face similar case-mix issues. Standardizing O/Es to the population of interest enables within system determination of outlier status among sites.

Similarly, it is difficult to directly compare VA risk-adjusted rates to those of non-VA systems. We intentionally avoided direct comparison to HCUP rates given differences including previously noted case-mix and weighting differences plus use of calendar years versus FY. Our methods enable valid trend comparisons between systems since they are not affected by case-mix differences. However, they do not account for different starting points between systems such that one may have more room to improve than another; further underlying case-mix may make changes more difficult to effect in one system versus another.

The relative insensitivity of the procedure-related mortality indicators to detect temporal changes or site differences in the VA may be due to the success of longstanding VA programs such as NSQIP, or because of inadequate sample sizes, (eg, esophageal cancer resection had only 0–12 cases in a given year). Other systems may face similar size sample issues in comparing procedure rates.

Other IQI-associated general limitations include the fact they measure outcomes, ie, mortality or utilization rates. As the end result of care, they are vitally important to measure, however, they are less immediately actionable than process measures. They also use administrative data. Whereas documentation of death and major procedures should be fairly complete,[45,46] accurate risk-adjustment requires accurate coding of comorbidities which may vary within and across sites.

Nevertheless, such indicators complement Healthcare Effectiveness Data and Information Set and Joint Commission measures already in place in most healthcare systems. In the VA, these indicators can supplement national QI programs, such as EPRP, which tracks various performance measures (primarily process measures and some intermediate outcomes), and NSQIP, which tracks major surgery outcomes,[12] thereby providing a more comprehensive picture of the quality of VA care. For example, EPRP tracks processes of care for patients admitted with AMIs such as percentage of patients prescribed aspirin within 24 hours of hospital arrival.[47] We would expect performance on this measure to be related to an IQI such as in-hospital AMI mortality rate.[48]

Although the IQIs are screens as opposed to definitive measures, successful QI methods used by NSQIP participants could be replicated using IQIs.[11] NSQIP distributes annual reports to each facility that show how they compare with other facilities. It includes mortality and aggregate morbidity

TABLE 2. Comparison of NIS Versus National VA Rates

Mortality Indicators*	System	Slope Estimate	Standard Error	Estimate Difference	*T*
Medical condition-related					
Acute myocardial infarction (without transfer) (IQI 15)†‡	NIS	−5.07	0.08	5.31	3.52§
	VA	−10.39	1.50		
Congestive heart failure (IQI 16)†	NIS	−3.84	0.11	1.14	1.82
	VA	−4.98	0.62		
Acute stroke (IQI 17)†	NIS	−5.25	0.31	1.61	2.32§
	VA	−6.86	0.62		
Gastrointestinal hemorrhage (IQI 18)	NIS	−2.81	0.21	0.97	1.19
	VA	−3.78	0.79		
Hip fracture (IQI 19)†	NIS	−2.74	0.09	1.55	5.02§
	VA	−4.29	0.29		
Pneumonia (IQI 20)	NIS	−4.89	0.17	2.81	3.32§
	VA	−7.70	0.83		
Procedure-related					
Esophageal cancer resection (IQI 8)†	NIS	−7.69	3.22	−14.31	1.09
	VA	6.62	12.75		
Pancreatic cancer resection (IQI 9)†	NIS	−6.67	1.20	6.87	0.50
	VA	−13.54	13.58		
Abdominal aortic aneurysm repair (IQI 11)†	NIS	−2.81	0.80	5.93	1.44
	VA	−8.74	4.03		
Coronary artery bypass graft (IQI 12)	NIS	−2.93	0.28	2.16	1.87
	VA	−5.09	1.12		
Craniotomy (IQI 13)	NIS	−4.11	0.59	0.01	0.00
	VA	−4.12	1.95		
Hip replacement (IQI 14)	NIS	−0.16	0.07	0.37	2.79§
	VA	−0.53	0.11		
PTCA (IQI 30)	NIS	−0.35	0.05	0.86	1.71
	VA	−1.21	0.50		
Carotid endarterectomy (IQI 31)	NIS	−0.60	0.06	−0.38	1.32
	VA	−0.22	0.28		
Utilization indicators					
Laparoscopic cholecystectomy (IQI 23)	NIS	13.49	1.17	−12.23	5.26§
	VA	25.72	2.01		
Incidental appendectomy (IQI 24)†	NIS	−0.57	0.16	0.79	3.03§
	VA	−1.36	0.21		
Bilateral cardiac catheterization (IQI 25)†	NIS	−7.25	0.96	−4.35	2.44§
	VA	−2.90	1.50		

*Rates are deaths per 100 discharges with the principal diagnosis or procedure unless otherwise specified. See online Appendix 1 for full definitions of indicators.

†Endorsed by the National Quality Forum.

‡Because we found similar results for AMI mortality with and without transfers (IQI 32), we present data for AMI with transfers only.

§$t > 1.96$ = significantly different slopes. A positive estimate difference indicates VA risk-adjusted rates are decreasing more rapidly than NIS rates.

NIS indicates National Inpatient Sample; CABG, coronary artery bypass grafting; PTCA, percutaneous transluminal coronary angioplasty; CEA, carotid endarterectomy; IQI, Inpatient Quality Indicators.

O/Es for all noncardiac surgery and separately for 8 subspecialities.[49] In 1996 the Salt Lake VA hospital was identified as a high outlier for general surgical morbidity.[50] After receiving this information, local clinicians reviewed charts of NSQIP-identified patients experiencing complications to determine practice patterns and identify possible provider issues. After finding that many wound complications resulted from contaminated wound closure, they developed and instituted wound infection and disruption prevention protocols which resulted in a clinically significant decrease in wound complications.[50]

With respect to the IQIs, we found considerable VISN-level variation in bilateral cardiac catheterization rates. One possible QI scenario would be as follows: management at a high outlier VISN examines rates within its facilities to determine whether any facility had higher than expected rates. If so, that facility's QI team would review charts of a sample of its cases to determine why patients underwent

bilateral rather than left heart catheterizations (eg, operator preference, patient characteristics). Literature review of best practices, and, although not used in the previous case, review of low outlier facilities systems and practices could be used to derive recommendations that would be implemented and their effect tested.

In summary, our findings suggest the AHRQ IQIs are useful for tracking and comparing outcomes representing potential quality of care issues within the VA. Such measurement, along with internal benchmarking, will help focus QI efforts on indicators showing the greatest intersite variation. Detailed case review of systems and processes at both high and low outlier sites, may identify failures amenable to improvement and successes that can be adopted by other sites respectively, with the ultimate goal of improving care across the system. Although population differences limit external benchmarking using current IQI risk-adjustment methods, trends over time may still be compared across populations and healthcare systems. Future research should entail improving risk adjustment to facilitate external benchmarking and identifying high risk populations to target for QI interventions.

REFERENCES

1. Institute of Medicine. Committee on the National Quality Report on Health Care Delivery. In: Hurtado MP, Swift EK, Corrigan JM, eds. *Envisioning the National Health Care Quality Report.* Washington, DC: National Academy Press; 2001.
2. Agency for Healthcare Quality and Research. Guide to inpatient quality indicators: quality of care in hospitals—volume, mortality, and utilization [version 3.0.]. Rockville, MD; Agency for Healthcare Quality and Research; 2006.
3. Guidance for using the AHRQ. Quality indicators for hospital-level public reporting of payment. Rockville, MD; Agency for Healthcare Quality and Research; 2004. AHRQ Pub. No. 04–0086-EF.
4. Fraser I. AHRQ quality indicators. Paper presented at: AHRQ QI User Meeting; September 27, 2005; Rockville, MD.
5. Colorado Health and Hospital Association: Colorado Hospital Quality. Available at: http://hospitalquality.org. Accessed November 24, 2005.
6. Agency for Healthcare Quality and Research. State Snapshots 2008. From the National Healthcare Quality Report. Available at: http://statesnapshots.ahrq.gov/snaps08/map.jsp?menuId=2&state=. Accessed October 25, 2009.
7. Centers for Medicare & 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.
8. National Quality Forum. NQF Endorsed Standards. Available at: http://www.qualityforum.org/Measures_List.aspx. Accessed June 1, 2009.
9. Centers for Medicare & Medicaid Services. Details for: Quality Measures for Reporting in Fiscal Year 2009 for 2010 Update. Available at: http://www.cms.hhs.gov/apps/media/zpress/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 February 15, 2010.
10. Jha AK, Perlin JB, Kizer KW, et al. Effect of the transformation of the Veterans Affairs Health Care System on the quality of care. *N Engl J Med*. 2003;348:2218–2227.
11. Khuri SF. The NSQIP: a new frontier in surgery. *Surgery*. 2005;138:837–843.
12. Demakis JG, McQueen L, Kizer KW, et al. Quality Enhancement Research Initiative (QUERI): a collaboration between research and clinical practice. *Med Care*. 2000;38(suppl 1):I17–I25.
13. Andrews RM, Russo A, Pancholi MS. Trends in hospital risk-adjusted mortality for select diagnoses and procedures, 1994–2004. Rockville, MD: Agency for Healthcare Research and Quality; 2007.
14. 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=F45815BFCDCE8C28&Form=MAINSEL&JS=Y&Action=%3E%3ENext%3E%3E&_MAINSEL=AHRQ%20Quality%20Indicators. Accessed December 1, 2009.
15. Subramanian U, Weinberger M, Eckert GJ, et al. Geographic variation in health care utilization and outcomes in veterans with acute myocardial infarction. *J Gen Intern Med*. 2002;17:604–611.
16. Tseng CL, Rajan M, Miller DR, et al. Use of administrative data to risk adjust amputation rates in a national cohort of Medicare-enrolled veterans with diabetes. *Med Care*. 2005;43:88–92.
17. Veterans Affairs 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, 2008.
18. Veterans Affairs Information Resource Center. VIReC Research User Guide: FY2002 VHA Medical SAS Outpatient Datasets. Available at: http://www.virec.research.med.va.gov/References/RUG/RUG-Outpatient02.pdf. Accessed December 1, 2008.
19. Davies SM, Geppert J, McClellan MD, et al. Refinement of the HCUP Quality Indicators. Technical Review 4. Rockville, MD: Agency for Health Care Research and Quality; May 2001. AHRQ Pub. No. 01–0035 (Prepared by the UCSF-Stanford Evidence-Based Practice Center under Contract # 290–97–0013).
20. Birkmeyer JD, Siewers AE, Finlayson EV, et al. Hospital volume and surgical mortality in the United States. *N Engl J Med*. 2002;346:1128–1137.
21. Hannan EL, Popp AJ, Feustel P, et al. Association of surgical specialty and processes of care with patient outcomes for carotid endarterectomy. *Stroke*. 2001;32:2890–2897.
22. Hannan EL, Wu C, Ryan TJ, et al. Do hospitals and surgeons with higher coronary artery bypass graft surgery volumes still have lower risk-adjusted mortality rates? *Circulation*. 2003;108:795–801.
23. Solomon RA, Mayer SA, Tarmey JJ. Relationship between the volume of craniotomies for cerebral aneurysm performed at New York state hospitals and in-hospital mortality. *Stroke*. 1996;27:13–17.
24. Meehan TP, Hennen J, Radford MJ, et al. Process and outcome of care for acute myocardial infarction among Medicare beneficiaries in Connecticut: a quality improvement demonstration project. *Ann Intern Med*. 1995;122:928–936.
25. Meehan TP, Fine MJ, Krumholz HM, et al. Quality of care, process, and outcomes in elderly patients with pneumonia. *JAMA*. 1997;278:2080–2084.
26. Warren JL, Penberthy LT, Addiss DG, et al. Appendectomy incidental to cholecystectomy among elderly Medicare beneficiaries. *Surg Gynecol Obstet*. 1993;177:288–294.
27. Malone ML, Bajwa TK, Battiola RJ, et al. Variation among cardiologists in the utilization of right heart catheterization at time of coronary angiography. *Cathet Cardiovasc Diagn*. 1996;37:125–130.
28. Guidelines for the clinical application of laparoscopic biliary tract surgery. *Surg Endosc.* 1998;12:191–192.
29. Agency for Healthcare Research and Quality. *Inpatient Quality Indicators: SAS Software Documentation* [computer program]. Version 3.1, Rev. 1. Rockvile, MD: Agency for Healthcare Research and Quality; 2007.
30. Rivard P, Elwy AR, Loveland S, et al. Applying patient safety indicators (PSIs) across healthcare systems: achieving data comparability. Vol 2. Rockville, MD: Agency for Healthcare Research and Quality; 2005:7–25. AHRQ Pub. No. 005–0021-CD.
31. AHRQ Quality Indicators 3M APR-DRG V24 Limited License Grouper SAS Documentation. Available at: http://www.qualityindicators.ahrq.gov/software.htm. Accessed September 1, 2009.
32. *AHRQ Quality Indicators Inpatient Quality Indicator Composite (March, 2007) SAS Documentation* [computer program]. Version 3.1a. Available at: http://www.qualityindicators.ahrq.gov/downloads/iqi/IQI_Composite_SAS_Documentation_v31.pdf. Accessed September 1, 2009.

33. 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.
34. US Census Bureau Regions. Available at: http://www.census.gov/geo/www/us_regdiv.pdf. Accessed July 15, 2009.
35. Baker DW, Einstadter D, Thomas CL, et al. Mortality trends during a program that publicly reported hospital performance. *Med Care*. 2002; 40:879–890.
36. Marciniak TA, Ellerbeck EF, Radford MJ, et al. Improving the quality of care for Medicare patients with acute myocardial infarction: results from the Cooperative Cardiovascular Project. *JAMA*. 1998;279:1351–1357.
37. Centers for Medicare and Medicaid Services. Medicare Quality Monitoring System (MQMS). Available at: http://www.cms.hhs.gov/QualityInitiativesGenInfo/15_MQMS.asp. Accessed May 1, 2009.
38. Volpp KG, Rosen AK, Rosenbaum PR, et al. Mortality among patients in VA hospitals in the first 2 years following ACGME resident duty hour reform. *JAMA*. 2007;298:984–992.
39. Volpp KG, Rosen AK, Rosenbaum PR, et al. Mortality among hospitalized Medicare beneficiaries in the first 2 years following ACGME resident duty hour reform. *JAMA*. 2007;298:975–983.
40. Asch SM, McGlynn EA, Hogan MM, et al. Comparison of quality of care for patients in the Veterans Health Administration and patients in a national sample. *Ann Intern Med*. 2004;141:938–945.
41. Lanska DJ, Kryscio R. Geographic distribution of hospitalization rates, case fatality, and mortality from stroke in the United States. *Neurology*. 1994;44:1541–1550.
42. Krumholz HM, Chen J, Rathore SS, et al. Regional variation in the treatment and outcomes of myocardial infarction: investigating New England's advantage. *Am Heart J*. 2003;146:242–249.
43. Cohen ME, Dimick JB, Bilimoria KY, et al. Risk adjustment in the American College of Surgeons National Surgical Quality Improvement Program: a comparison of logistic versus hierarchical modeling. *J Am Coll Surg*. 2009;209:687–693.
44. Agha Z, Lofgren RP, VanRuiswyk JV, et al. Are patients at Veterans Affairs medical centers sicker? A comparative analysis of health status and medical resource use. *Arch Intern Med*. 2000;160:3252–3257.
45. Kashner TM. Agreement between administrative files and written medical records: a case of the Department of Veterans Affairs. *Med Care*. 1998;36:1324–1336.
46. Sohn MW, Arnold N, Maynard C, et al. Accuracy and completeness of mortality data in the Department of Veterans Affairs. *Popul Health Metr*. 2006;4:2.
47. Office of Quality and Performance. FY 2009, Q3 Technical Manual for the VHA Performance Measurement System. April 1, 2009.
48. Erne P, Radovanovic D, Urban P, et al. Early drug therapy and in-hospital mortality following acute myocardial infarction. *Heart Drug*. 2003;3:134–140.
49. Khuri SF, Daley J, Henderson W, et al. The Department of Veterans Affairs' NSQIP: the first national, validated, outcome-based, risk-adjusted, and peer-controlled program for the measurement and enhancement of the quality of surgical care. National VA Surgical Quality Improvement Program. *Ann Surg*. 1998;228:491–507.
50. Neumayer L, Mastin M, Vanderhoof L, et al. Using the Veterans Administration National Surgical Quality Improvement Program to improve patient outcomes. *J Surg Res*. 2000;88:58–61.

JHH Open

Journal of Human Hypertension (2010) 24, 9–18

www.nature.com/jhh

ORIGINAL ARTICLE

The epidemiology and management of severe hypertension

AM Borzecki[1,2,3], B Kader[1,2] and DR Berlowitz[1,2,3]

[1]*Center for Health Quality Outcomes and Economic Research, Bedford VAMC, Bedford, MA, USA;* [2]*Department of Health Policy and Management, Boston University School of Public Health, Boston, MA, USA and* [3]*Section of General Internal Medicine, Boston University School of Medicine, Boston, MA, USA*

Hypertension guidelines stress that patients with severe hypertension (systolic blood pressure (BP) ⩾180 or diastolic BP ⩾110 mm Hg) require multiple drugs to achieve control and should have close follow-up to prevent adverse outcomes. However, little is known about the epidemiology or actual management of these patients. We retrospectively studied 59 207 veterans with hypertension. Patients were categorized based on their highest average BP over an 18-month period (1 July 1999 to 31 December 2000) as controlled (<140/90 mm Hg), mild (140–159/90–99 mm Hg), moderate (160–179/100–109 mm Hg) and severe hypertension. We examined severe hypertension prevalence, pattern, duration, associated patient characteristics, time to subsequent visit, percentage of visits with a medication increase, and final BP control and antihypertensive medication adequacy. Twenty-three per cent had ⩾1 visit with severe hypertension, 42% of whom had at least two such visits; median day with severe hypertension was 80 (range 1–548). These subjects were significantly older, more likely black, and with more comorbidities than other hypertension subjects. Medication increases occurred at 20% of visits with mild hypertension compared to 40% with severe hypertension; $P<0.05$). At study end, 76% of patients with severe hypertension remained uncontrolled; severe hypertension subjects with uncontrolled BP were less likely to be on adequate therapy than those with controlled BP (43.7 vs 45.4%). Among hypertensive veterans, severe hypertension episodes are common. Many subjects had relatively prolonged elevations, with older, sicker subjects at highest risk. Although, follow-up times are shorter and antihypertensive medication use greater in severe hypertension subjects, they are still not being managed aggressively enough. Interventions to improve providers' management of these high-risk patients are needed.

Journal of Human Hypertension (2010) **24,** 9–18; doi:10.1038/jhh.2009.37; published online 14 May 2009

Keywords: epidemiology; chronic hypertension; therapy

Introduction

Hypertension is among the most prevalent chronic conditions worldwide; with rates as high as 70% among adults in developed countries such as Poland.[1] Although hypertension is usually asymptomatic, it may be associated with considerable morbidity and mortality. The higher the blood pressure (BP), the greater the risk for adverse outcomes including development of coronary artery disease, congestive heart failure, stroke and kidney disease.[2] Hypertension treatment has been clearly shown to reduce this risk.[2–4]

Accordingly, the current World Health Organization/International Society of Hypertension and the European Society of Hypertension guidelines and the prior Joint National Committee on Prevention, Detection, Evaluation and Treatment of High Blood Pressure (JNC) guidelines classify BP into grades or stages based on the absolute BP level.[5–7] Although absolute cardiovascular risk is based not only on BP levels, but associated cardiovascular risk factors or target organ damage, individuals with the highest levels, grade/stage[3] or severe hypertension (that is, systolic BP ⩾180 mm Hg or diastolic BP ⩾110 mm Hg)[5,8] have a 20–30% 10-year risk of cardiovascular disease, that increases to very high risk, >30%, in the presence of any risk factors or target organ damage.[5] Further, these subjects are also at high short-term risk for serious cardiovascular events, the risk increasing with the degree and speed of elevation. Because of this, guidelines also stress that such patients should have close follow-up with reassessment at most within 1 week, and will require multiple drugs to achieve control.[5,9,10]

Although much has been written about the epidemiology and management of the general hypertension population, relatively little is known about these issues in those with severe hypertension. Limited cross-sectional data suggest a prevalence among those with hypertension in the

Correspondence: Dr AM Borzecki, Center for Health Quality Outcomes and Economic Research, Bedford VA Hospital (152), 200 Springs Road, Bedford, MA, 1730, USA.
E-mail: amb@bu.edu
Received 8 October 2008; revised 10 April 2009; accepted 11 April 2009; published online 14 May 2009

8–19% range.[10–13] How long such patients are exposed to high BPs, or how they are being managed in actual practice is relatively unknown, with existing data based on small samples and select populations.[12–14]

The purpose of the present study was to examine the following issues with respect to patients with severe hypertension: (1) what is the epidemiology of severe hypertension, in terms of prevalence, pattern of severe hypertension and duration of severe hypertension? (2) what patient characteristics are associated with severe hypertension? (3) how are patients with severe hypertension being managed in everyday practice, including time to next visit, percentage of visits with a medication increase and number of BP medications by final BP control?

Materials and methods

Study population

The study population is previously described.[15] Briefly, we identified individuals with hypertension who were receiving regular outpatient care at geographically diverse sites within the largest integrated health-care system in the United States, the Veterans Health Administration (VA), using the VA's National Patient Care Database (NCPD) through 2000, eligible subjects: (1) had at least two NPCD-listed hypertension diagnoses, ICD-9-CM code 401, between 1 July 1997 and 30 June 1999; and (2) were regular VA users (that is, $\geqslant 3$ NPCD-listed visits to a general medicine or subspecialty medical clinic between 1 July 1999 and 31 December 2000). Subjects were followed from 1 July 1999 through 31 December 2000. The study protocol was approved by the Bedford VA Hospital's institutional review board.

Data collection

Data sources were the Veterans Health Information Systems and Technology Architecture (VISTA), that is, the VA's electronic record system, the NPCD and Medicare files. VISTA, maintained at each site, contains multiple files including clinical data such as vital signs, laboratory results, pharmacy records and provider notes. (We previously found the vitals file very complete with provider notes contributing minimal additional BP information.)[16] VISTA also contains diagnoses and procedure information from all outpatient visits and inpatient stays, which is transferred to a central data repository and incorporated into the NPCD.[17]

Demographics and comorbidities were obtained from the NPCD (1998 to 2000 inclusive), supplemented by Medicare denominator, inpatient and outpatient files (MEDPAR, Carrier and Outpatient files). We used the Medicare denominator file as our primary race source because this is patient identified and supplemented this with VA race data. Baseline comorbidities were identified by the presence of specified ICD-9-CM codes in the 12 months preceding and including the index visit.[15] Height, weight and BP were obtained from VISTA vitals, and medications from VISTA pharmacy files (1 July 1999 to 31 December 2000).

Analyses

Available BPs were averaged at each visit. All visits during 1 July 1999 through 31 December 2000 were categorized according to the average systolic and diastolic BPs as follows: controlled ($<$140/90 mm Hg), mild (140–159/90–99 mm Hg), moderate (160–179/100–109 mm Hg) and severe ($\geqslant$180/110 mm Hg) hypertension. Patients were then grouped based on their visit with the highest average BP category unless otherwise specified. If systolic and diastolic BPs fell into different categories, the higher category was used. Among patients with more than one severe hypertension episode, patient level analyses considered only the first episode unless otherwise indicated.

Objective 1: examine the epidemiology of patients with severe hypertension. We first determined the prevalence of severe hypertension in our cohort based on the number of subjects who had at least one visit with severe hypertension divided by the total sample of hypertension patients. We also looked at the proportion of all visits with severe hypertension and the median number of visits with severe hypertension per patient. Next, we characterized the pattern of severe hypertension by determining the following proportions: (1) those who had a single visit with severe hypertension, with no severe hypertension at the next visit (that is, an isolated spike in their BP); (2) those who had severe hypertension at two or more consecutive visits (persistent severe hypertension); (3) those who had an intervening visit with a lower BP but then another visit with severe hypertension (recurrent severe hypertension). We further determined the median number of recurrences per patient. In addition, we examined time in days to a subsequent visit with a BP $<$180/110 mm Hg as a proxy for the duration of (that is, exposure to) severe hypertension, and compared among severe hypertension groups. We summed these time periods for patients with recurrent episodes. (This was examined first by censoring subjects with no visit after the severe hypertension visit at either the last day of the study period or death, and then by excluding them.)

Objective 2: patient characteristics associated with severe hypertension. We next compared patients with severe hypertension to subjects without severe hypertension, grouped by highest BP category, with respect to age, gender, race, baseline comorbidities (including cardiovascular risk factors and pre-existing cardiovascular conditions, and Charlson

index),[18,19] body mass index (BMI). We also compared severe hypertension subgroups with respect to these same characteristics.

We used χ^2-analysis for categorical variables, and *t*-tests, one-way analysis of variance (ANOVA) with Tukey's multiple comparison procedure, or Wilcoxon rank-sum tests as appropriate for continuous variables.

As part of a sensitivity analysis, we repeated objectives 1 and 2 assigning subjects to BP categories based on the first visit of the study period.

Objective 3: examine the management of patients with severe hypertension. We next examined actions potentially under clinician control. The following analyses were carried out at the visit level: (1) we determined the time in days to the next visit that included a BP assessment following a visit with severe hypertension. We used Wilcoxon rank-sum tests to compare the median interval to the next visit among severe hypertension subgroups, and between visits with severe hypertension to visits with moderate hypertension, mild hypertension or controlled BP. (2) We then examined the percentage of visits associated with a medication increase. To determine medication increases at a visit, we defined baseline medication use based on the 6-month period preceding study entry (that is, we examined medication data out to 1 January 1999); using methods we previously developed, a patient was considered to have an increase at a given visit if the dose of an existing medication was increased or a new medication was started.[20,21] We compared the proportion of visits with a medication increase among the severe hypertension subgroups then compared among visits with moderate, mild hypertension or controlled BP using χ^2-tests. We also looked at the impact of this increase on BP control at the subsequent visit for the severe hypertension group.

Although we lacked information on specialty referrals for BP management or to exclude secondary hypertension, as a proxy, we examined differences in proportions of patients seen in subspecialty clinics at which BP management might be addressed by χ^2-analysis and generating 95% confidence intervals. These clinics included hypertension, renal, cardiology and endocrinology clinics. We further looked at the frequency of these visits with a primary diagnoses related to hypertension (ICD-9-CM codes 401, 402, 403, 404 and 405 (secondary hypertension)) and the frequency of any 405 code at these visits.

To examine medication use and BP control at study end, we then categorized subjects based on their highest average BP at a visit before the last study visit. We compared the percentage of patients who achieved a BP $<140/90$ mm Hg at the last study visit by BP group. We next compared BP groups by mean number of prescribed antihypertensive medications (using ANOVA), percentage of subjects on each major class of antihypertensive medication and the percentage on adequate therapy defined as a regimen containing at least three different classes of drugs at least one of which had to be a loop or thiazide diuretic[9] at the last study visit using χ^2-tests. We then examined the mean number of prescribed antihypertensive medications at the time of the last visit by prior highest BP category and final BP control (that is, BP $<140/90$ mm Hg; yes/no) using linear regression models, comparing both within BP groups, and across groups for controlled vs uncontrolled. We also repeated analyses comparing among severe hypertension subgroups. Finally, we examined adequacy of antihypertensive therapy at the last study visit and final BP control. Within each BP group defined by the highest BP before the last study visit, we used χ^2-tests to examine likelihood of adequate therapy at study end by final BP control (yes/no); we also compared adequacy across BP categories. We performed similar comparisons among severe hypertension groups.

Results

Objective 1

Our total sample consisted of 59 207 subjects. The mean age was 65 years: 58% were white and 97% were men.[15] Twenty-three per cent ($N=13\,735$) had at least one visit with severe hypertension; among these subjects the median number of such visits was 1 (range 1–57). Six per cent of all visits (21 992/325 105) were associated with severe hypertension. Of note, 87% of visits with a BP of $\geqslant 180/110$ mm Hg had only one BP documented on the day of the visit.

Of the severe hypertension group, 18% had severe hypertension at two or more consecutive visits (persistent), 23% had one or more intervening visits with a lower BP but then another visit with severe hypertension (recurrent); 58% had a single visit with a BP $\geqslant 180/110$ mm Hg (isolated spike). In 17% of subjects with an isolated spike, there was no subsequent documented visit. The median number of recurrences among the severe hypertension group was 0, range 0–34 (interquartile range 0–1). The median number of days with severe hypertension overall was 80 (range 1–548; Table 1) and varied

Table 1 Total exposure to severe hypertension

BP category	*Exposure to severe hypertension Days, median (range)*[a]
Severe hypertension, overall	80 (1–548)
Persistent	182 (2–548)
Recurrent	103 (2–534)
Isolated spike	50 (1–549)

Abbreviation: BP, blood pressure.
[a]Significant difference with respect to number of days exposed between groups ($P<0.001$).

Table 2 Baseline patient characteristics by highest BP group

Characteristic	*Controlled* N = *5170*	*Mild* N = *19 290*	*Moderate* N = *21 012*	*Severe* N = *13 735*
Age, mean (s.d.), years	63.7 (11.3)	63.8 (11.2)	65.5 (10.9)	67.0 (10.8)[a]
Gender, female, no. (%)	121 (2.3)	516 (2.7)	604 (2.9)	453 (3.3)[a]
Race, no. (%), white	3988 (77.1)	14 563 (75.5)	15 573 (74.1)	9545 (69.5)[a]
Black	659 (12.8)	2,842 (14.7)	3,741 (17.8)	3302 (24.0)[a]
Hispanic	71 (1.4)	233 (1.2)	300 (1.4)	153 (1.1)
Others	108 (2.1)	358 (1.9)	341 (1.6)	221 (1.6)
Unknown	344 (6.7)	1294 (6.7)	1057 (5.0)	514 (3.7)[a]
BMI, mean (s.d.), $kg\,m^{-2}$	28.8 (5.4)	29.4 (5.5)	29.5 (5.8)	29.2 (6.1)[a]
Comorbidities				
Cerebrovascular disease, no. (%)	669 (12.9)	2261 (11.7)	2781 (13.2)	2305 (16.8)[a]
Congestive heart failure, no. (%)	1055 (20.4)	2338 (12.1)	2354 (11.2)	1814 (13.2)[a]
Coronary artery disease, no. (%)	2392 (46.3)	6998 (36.3)	7128 (33.9)	4802 (35.0)[a]
Diabetes, no. (%)	1584 (30.6)	5841 (30.3)	6996 (33.3)	5263 (38.3)[a]
Hyperlipidaemia, no. (%)	2563 (49.6)	8927 (46.3)	8895 (42.3)	5151 (37.5)[a]
Peripheral vascular disease, no. (%)	532 (10.3)	1688 (8.8)	2141 (10.2)	1760 (12.8)[a]
Renal disease, no. (%)	340 (6.6)	1050 (5.4)	1378 (6.6)	1430 (10.4)[a]
Tobacco use, no. (%)	503 (9.7)	1897 (9.8)	1951 (9.3)	1219 (8.9)[a]
Charlson index, mean (s.d.)	1.3 (1.4)	1.2 (1.2)	1.2 (1.2)	1.4 (1.3)[a]

Abbreviations: BMI, body mass index; s.d., standard deviation.
[a]$P<0.05$ for among-group comparisons and pair-wise comparisons; the severe hypertension group was significantly different than the three other hypertension groups with respect to all the characteristics listed. However, with respect to race, the prevalence of Hispanics in the severe hypertension group was only significantly different compared to the moderate hypertension group.

significantly by severe hypertension subgroup. (Overall exposure did not change significantly if we excluded subjects with a single visit.)

Categorizing subjects using the first study period visit, 7% of subjects had severe hypertension, of whom 27% had persistent, 28% had recurrent and 45% had an isolated spike. Exposure trends were similar.

Objective 2

Compared to subjects with lower BPs, severe hypertension subjects were older, ($67.0 + 10.8$ vs $64.6 + 11.2$ years; $P<0.05$), were more likely to be black (24 vs 16%; $P<0.05$) and women (3.3 vs 2.7%, $P<0.05$) compared to all other subjects combined ($N = 45\,472$). They also had more total comorbidities (for example, Charlson index $1.4 + 1.3$ vs $1.2 + 1.2$, $P<0.05$) and a higher prevalence of specific comorbidities such as diabetes, renal disease and peripheral vascular disease; although when further divided into BP categories, the controlled BP group had the highest prevalence of coronary artery disease, congestive heart failure and hyperlipidaemia. (See Table 2 for comparisons among BP categories.)

Among severe hypertension subgroups, the recurrent and persistent were similar with respect to age and race distribution, but the recurrent had the most comorbidities followed by the persistent then the isolated spike group. The isolated spike group was younger and more likely to be white (see Table 3). Similar trends were seen when subjects were categorized by the first study visit (data not shown; available from authors).

Objective 3

The median time to a subsequent visit after presentation was 42 days (range 1–503) for severe hypertension, 52 days (1–513) for controlled BP, 60 days (1–518) for mild and 55 (1–496) for moderate hypertension visits ($P<0.05$ for between-group differences).

Hypertension medication increases occurred at 40% of severe hypertension, 10% of controlled, 20% of mild and 32% of moderate hypertension visits ($P<0.05$). Among visits with severe hypertension, medications were increased at 41% of visits with severe hypertension among the persistent group, 36% of the recurrent group and 42% of the isolated spike group ($P<0.05$). If a medication was increased at a given visit with severe hypertension, BP was controlled 13.2% of the time at the next visit, versus 12.5% of next visits if there was no medication increase. (This difference was not significant (NS)).

The proportion of patients seen in a subspecialty clinic at which BP management might be addressed not surprisingly tended to increase with increasing BP category, and was significantly higher in the severe hypertension group (except for cardiology clinic visits) presumably in part driven by the higher prevalence of renal disease or diabetes in the more severe group (Table 4). The frequency of such visits associated with a primary diagnosis of hypertension ranged from 6.2% among the controlled group, 8.9% of the mild, 13.3% of the moderate, to 20.0% among the severe hypertension group. In severe hypertension subgroup, this ranged from 17.5% among the isolated spike group, 18.0% among the recurrent group and 29.8% among the persistent group ($P<0.5$

Table 3 Baseline patient characteristics by severe hypertension group

Characteristic	*Persistent severe* N = *2517*	*Recurrent severe* N = *3210*	*Isolated spike* N = *8008*
Age, mean (s.d.), years	67.4 (10.8)	67.7 (10.6)	66.6 (10.9)[a]
Gender, female, no. (%)	80 (3.2)	104 (3.2)	269 (3.4) NS
Race, no. (%), white	1631 (64.8)	2144 (66.8)	5770 (72.1)[a]
Black	734 (29.2)	910 (28.4)	1658 (20.7)[a]
Hispanic	26 (1.0)	42 (1.3)	85 (1.1)
Others	40 (1.6)	40 (1.3)	141 (1.8)
Unknown	86 (3.4)	74 (2.3)	354 (4.4)[a]
BMI, mean (s.d.), $kg\,m^{-2}$	29.2 (6.1)	29.0 (6.1)	29.4 (6.1)[a]
Comorbidities			
Cerebrovascular disease, no. (%)	470 (18.7)	612 (19.1)	1223 (15.3)[a]
Congestive heart failure, no. (%)	350 (13.9)	490 (15.3)	974 (12.2)[a]
Coronary artery disease, no. (%)	834 (33.1)	1217 (37.9)	2751 (34.4)[a]
Diabetes, no. (%)	1064 (42.3)	1396 (43.5)	2803 (35.0)[a]
Hyperlipidaemia, no. (%)	912 (36.2)	1178 (36.7)	3061 (38.2)[a]
Peripheral vascular disease, no. (%)	323 (12.8)	514 (16.0)	923 (11.5)[a]
Renal disease, no. (%)	291 (11.6)	465 (14.5)	674 (8.4%)[a]
Tobacco use, no. (%)	211 (8.4%)	287 (8.9%)	721 (9.0) NS
Charlson index, mean (s.d.)	1.4 (1.3)	1.6 (1.4)	1.3 (1.3)[a]

Abbreviations: BMI, body mass index; NS, not significant; s.d., standard deviation.
[a]$P<0.05$ for among-group comparisons and pair-wise comparisons; recurrent and persistent groups were similar with respect to age, white and black race distribution, BMI, prevalence of cerebrovascular disease, congestive heart failure, diabetes and hyperlipidaemia.

Table 4 Percentage of patients with subspecialty clinic visits[a]

Highest BP category	*Hypertension clinic % (95% CI)*	*Renal clinic % (95% CI)*	*Cardiology clinic % (95% CI)*	*Endocrinology clinic % (95% CI)*	*Any subspecialty clinic*[b] *% (95% CI)*
Controlled ($N=5170$)	0.2 (0.1–0.4)	0.6 (0.4–0.8)	9.0 (8.2–9.8)	1.5 (1.2–1.7)	10.6 (9.8–11.5)
Mild ($N=19\,290$)	0.3 (0.2–0.4)	0.9 (0.8–1.1)	8.6 (8.2–9.0)	1.6 (1.4–1.8)	10.8 (10.3–11.2)
Moderate ($N=21\,012$)	0.4 (0.3–0.5)	1.7 (1.5–1.8)	8.6 (8.2–8.9)	1.9 (1.7–2.1)	11.5 (11.1–12.0)
Severe ($N=13\,735$)	0.8 (0.7–1.0)	3.7 (3.4–4.0)	8.9 (8.4–9.4)	2.2 (2.0–2.5)	14.1 (13.6–14.7)
Persistent ($N=2517$)	1.2 (0.9–1.7)	4.6 (3.8–5.5)	7.2 (6.3–8.3)	2.5 (1.9–3.1)	13.7 (12.4–15.1)
Recurrent ($N=3210$)	1.0 (0.7–1.4)	6.2 (5.4–7.1)	11.7 (10.6–12.8)	2.8 (2.3–3.4)	19.4 (18.0–15.1)
Isolate spike ($N=8008$)	0.7 (0.5–0.9)	2.4 (2.1–2.7)	8.3 (7.8–9.0)	2.0 (1.7–2.3)	12.2 (11.5–12.9)

Abbreviations: BP, blood pressure; CI, confidence interval.
Non-overlapping of CIs indicates the proportions (percentages) are significantly different from each other.
[a]Clinics at which hypertension is likely to be treated.
[b]Any subspecialty clinic: hypertension, renal, cardiology or endocrinology.

for between-group differences). The frequency of a code for secondary hypertension ranged from 0% among endocrinology visits, 0.1% for cardiology, 0.9% for renal, to 2.8% of all hypertension clinic visits.

When categorized by the highest available BP before the last visit of the study, at the end of study 24% of the severe hypertension group had controlled BP, compared to 73% of the controlled, 51% of mild and 33% of the moderate group ($P<0.05$). For the severe hypertension subgroups, the isolated spike group was most likely to be controlled (27%), followed by the recurrent (21%) and the persistent groups (16%; $P<0.05$ for between-group differences). (Of note, 1390 subjects with only one visit with a BP were excluded from this analysis; 500 had controlled, 525 had mild, 243 had moderate and 122 (9%) had severe hypertension. We found similar results to those reported below if we included these subjects in analyses and assumed they had an additional visit and their BP category was stable (data not shown; available from authors).)

At the end of study, the mean number of BP medications increased significantly with increasing BP category (Table 5; Figure 1). In addition, severe hypertension patients were significantly more likely to be on all classes of antihypertensives compared to the other groups (Table 5). Among severe hypertension subgroups, the persistent and recurrent subgroups were on more medications at study end than

Table 5 Medication use at study end by highest preceding BP category

Characteristic	*Controlled* N = *6428*	*Mild* N = *19 488*	*Moderate* N = *19 528*	*Severe* N = *12 373*	*Persistent* N = *2264*	*Recurrent* N = *2678*	*Isolated spike* N = *7431*
No. medications, mean (s.d.)	2.0 (1.1)	2.1 (1.1)	2.3 (1.2)	2.8 (1.4)[a]	3.2 (1.5)	*3.2 (1.5)*	2.6 (1.3)[a]
ACEIs/ARBs, no. (%)	3167 (49.3)	9525 (48.9)	10 193(52.2)	7121 (57.6)[a]	1400 (61.8)	1548 (57.8)	4173 (56.2)[a]
α-Blockers, no. (%)	1196 (18.6)	3952 (20.3)	4438 (22.7)	2980 (24.1)[a]	572 (25.3)	688 (25.7)	1720 (23.2)[a]
β-Blockers, no. (%)	2201 (34.2)	5797 (29.8)	5836 (29.9)	4494 (36.3)[a]	934 (41.3)	1056 (39.4)	2504 (33.7)[a]
CCBs, no. (%)	1912 (29.7)	6536 (34.1)	7765 (39.8)	5760 (46.6)[a]	1142 (50.4)	1333 (49.8)	3285 (44.2)[a]
Diuretics, no. (%)	2503 (38.9)	7595 (39.0)	8383 (42.9)	5980 (48.3)[a]	1161 (51.3)	1346 (50.3)	3473 (46.7)[a]
Other vasodilators, no. (%)[b]	173 (2.7)	548 (2.8)	938 (4.8)	1527 (12.3)[a]	421 (18.6)	479 (17.9)	627 (8.4)[a]
Adequate therapy, no. (%)[c]	1519 (23.6)	4608 (23.7)	6018 (30.8)	5457(44.1)[a]	1197 (52.9)	1383 (51.6)	2877 (38.7)[a]

Abbreviations: ACEIs, angiotensin-converting enzyme inhibitors; ARBs, angiotensin receptor blockers; BMI, body mass index; CCBs, calcium channel blockers; s.d., standard deviation.
[a]$P<0.05$ for among-group and pair-wise comparisons; analyses were carried out separately for the BP groups and severe hypertension subgroups. In pair-wise comparisons, for BP groups, the mean number of medications and the percentage on each medication class or on adequate therapy increased with increasing BP group, except for β-blocker use where the controlled group was more likely to be on a β-blocker than the mild or moderate group. Also, the controlled and mild groups were not significantly different with respect to ACEI/ARB use, diuretic use, other vasodilators use and percentage on adequate therapy. For the severe hypertension subgroups, the mean number of medications, the percentage on each medication class or adequate therapy was significantly lower for the isolated spike group than the persistent or recurrent, except for ACEI/ARB use which was similar to the recurrent. The persistent and recurrent groups were similar in all comparisons except for ACEI/ARB use.
[b]Includes clonidine, hydralazine, methyldopa and minoxidil.
[c]Adequate therapy: a regimen containing at least 3 different classes of drugs at least one of which is diuretic.

the isolated spike subgroup (Table 5; $P<0.05$ for between-group differences). The proportion on each medication class was significantly lower for the isolated spike group than the persistent or recurrent group, except for ACEI/ARB use which was similar to the recurrent group (Table 5). Despite a trend towards higher use of each medication class among the persistent group compared to the recurrent, the differences were not significant.

Across all BP groups, those with controlled BP at study end (<140/90 mm Hg) were on more medications than those with uncontrolled BP (although this difference was not significant for the severe hypertension group; Figure 1). Within severe hypertension subgroups, the mean number of medications at study end was only significantly different between controlled and uncontrolled subjects for the isolated spike group (Figure 1).

Use of three or more medications including a diuretic at the last visit increased by increasing preceding highest BP category (although the difference between the controlled and mild groups was not significant) (see Table 5 and Figure 2). Among the severe hypertension subgroups, those with persistent or recurrent elevations were more likely to be on adequate multi-drug therapy compared to the isolated spike group (Table 5). Similar to the medication class analysis, the proportions on adequate therapy in the persistent and recurrent groups were not significantly different (Table 5).

When examining adequacy of therapy and final BP control, subjects with uncontrolled BP at the last visit were less likely to be on adequate multi-drug therapy than those with controlled BP within any given BP group or severe hypertension subgroup (Figure 2), although differences were not significant for the severe hypertension group (45.4 vs 43.7%, NS; Figure 2). In subgroups, differences were only significant among the isolated spike subgroup (Figure 2).

Discussion

This is the largest study to date examining the epidemiology and management of patients with severe hypertension. This is also the first to characterize the pattern of severe hypertension, to examine management of such patients and to compare among hypertension categories. We found that severe hypertension is relatively common, with almost one quarter of subjects having at least one visit with severe hypertension. Further, in many cases this represented more than a single isolated spike, with over 40% having persistent or recurrent severe elevations.

Factors associated with severe hypertension were similar to those found in previous studies of poor BP control. Subjects with severe hypertension were older, more likely to be black, women, and had more medical comorbidities than subjects with lower BPs.

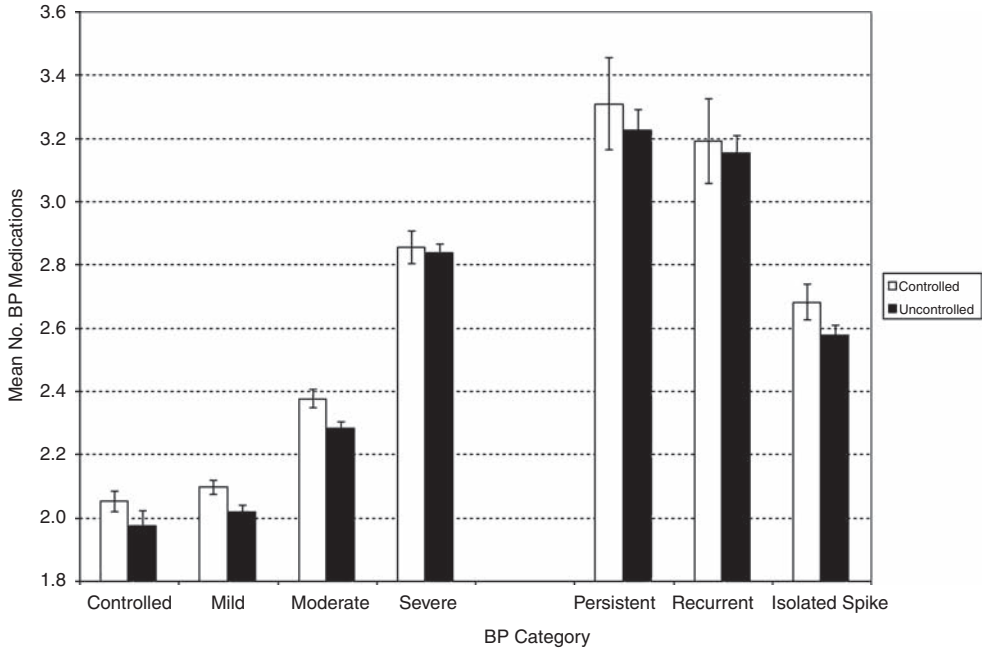


Figure 1 Mean number of antihypertensive medications at study end by highest preceding blood pressure (BP) category and final BP control. Means shown with surrounding bars representing 95% confidence intervals. (Error bars overlap for those in the controlled BP group with controlled vs uncontrolled BP at study end. However, by Student's *t*-test, the means of controlled vs uncontrolled are significantly different, $P<0.05$.)

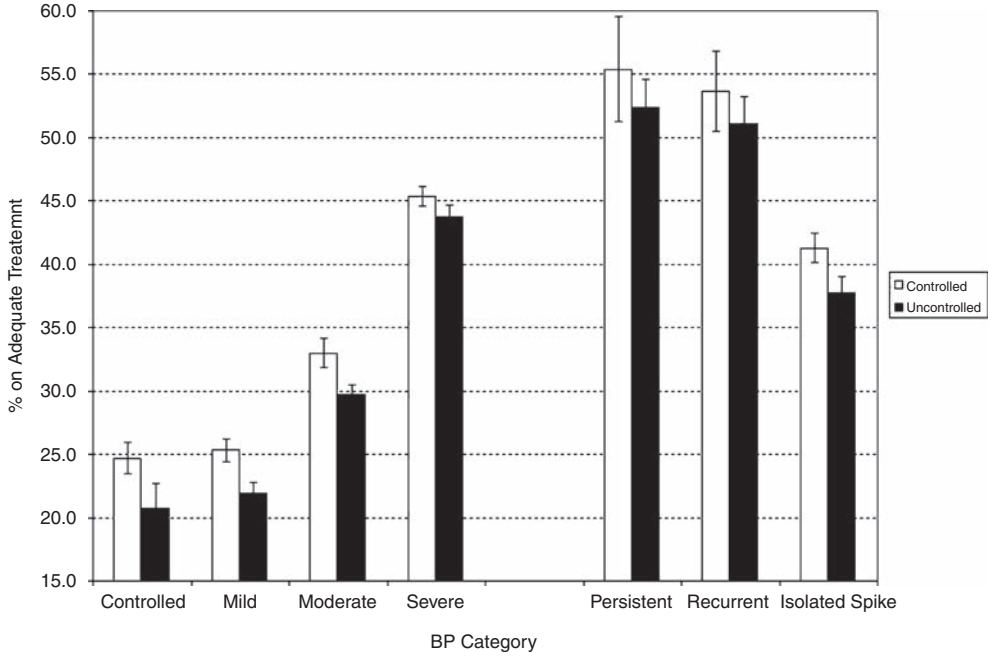


Figure 2 Adequacy of blood pressure (BP) treatment at study end by highest preceding BP category and final BP control. Bars represent 95% confidence intervals.

Similarly, among those with severe hypertension, those with recurrent or persistent elevations were older and women. The recurrent group had more medical comorbidities than those with persistent elevations or the isolated severe group. Analyses of National Health and Nutrition Examination Survey data likewise found age and being black were associated with poorer control.[22,23] An analysis of data from the Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial, additionally

found being woman, having diabetes and an elevated creatinine were associated with lack of control at follow-up.[24] We found similar associations in our previous examination of factors associated with poor BP control in this sample.[15]

We also found that patients with severe hypertension are often not treated aggressively enough. The median follow-up time after a visit with severe hypertension was 42 days, which is much longer than the guideline-recommended interval. Although medications were more likely to be increased as the severity of BP category increased, this still occurred at less than 50% of visits with severe hypertension and many subjects with severe hypertension remained uncontrolled at the end of the study. Although mean number of medications increased by BP category, within any category, including the severe hypertension group, subjects with uncontrolled BP at study end tended to be on fewer medications than controlled subjects. Further, a significant percentage of the severe hypertension group were not on adequate therapy at study end and continued to have poorly controlled BP. Among the severe hypertension subgroups, those on more medications tended to be better controlled at study end. Subjects with persistent severe hypertension were more likely to have their medications increased but were slightly less likely to be controlled compared to the recurrent group.

Existing data specifically examining severe hypertension epidemiology and management are very limited, being based on small populations and select samples. The largest observational study examined outcomes in 142 emergency room (ER) patients presenting with BPs $\geqslant$220 mm Hg systolic or $\geqslant$120 mm Hg diastolic.[12] Among those treated as urgencies and discharged from the ER, the mean time to a follow-up appointment was 21 days. Thirty percent of subjects returned to the ER with uncontrolled hypertension, 14% with hypertensive complications, within an average of 33 days. Drug management was also only examined in the acute ER setting. Longer-term use of drugs, BPs and clinical outcomes were not tracked. A cross-sectional Spanish study assessing cardiovascular risk and comorbidities in hypertension patients attending primary care practices found the following prevalences of BP categories among 1413 subjects: controlled 3%, mild 50%, moderate 39%, severe 8%.[13] The mean age of their sample was similar to ours at 65.3 (11.4) years. Among high-risk patients (based on comorbidities and absolute BP levels) there was no medication change in 30%. No information is given about the severe hypertension subjects as a distinct group. Only one study specifically examined subjects with severe hypertension.[12] Lalljie and Lalljie[12] studied management and BP outcomes in 48 subjects with severe hypertension of 252 (19%) patients presenting to a Jamaican hypertension specialty practice. Follow-up data were only reported on 31 of these subjects, of whom more than 50% achieved BP control during up to 2 years of follow-up with most requiring at least four drugs. Unlike our study, subjects were younger (62% were $<$65 of age), more likely to be women, with lower baseline comorbidity rates (31% had diabetes, 21% had heart failure and 15% had coronary artery disease).

This study has a few caveats. It was performed in a sample of predominantly male veterans with relatively high disease burden and good access to medical care and medications. Therefore, findings may not be generalizable to other settings. Our data predate VA quality performance data that show improvement in control rates over time, such that the current prevalence of severe hypertension is likely lower.[25] However, given the prevalence of hypertension in the VA population has increased from 37% in 1999 to 55–60% in 2006 (based on ICD-9 codes), in absolute numbers this still likely represents many patients with severe hypertension.[26] We lack data on some management aspects of severe hypertension subjects that may be of interest such as referrals to hypertension specialists (although we do know that less than 1.0% of these patients were seen in a hypertension clinic) or investigations to exclude secondary hypertension.

We have also demonstrated that the method used to define the BP group, for example, highest average BP at any visit vs the first visit of the 18-month study period produces dramatically different values with respect to severe hypertension prevalence. Further deviation from the true prevalence may result because BPs were obtained from the vitals file, with only one BP available for almost 90% of visits; we may be missing BPs present in provider notes. In addition, we cannot exclude white coat hypertension because we do not have access to home BP measures or 24 h ambulatory blood pressure monitoring. However, in a previous study, we found that provider notes contributed minimal additional BP information beyond the vitals file, including rare documentation of ambulatory or home BPs.[16]

Our examination of hypertension management was based on highest BP as opposed to the first study BP because we wanted to determine providers' responses to such high BPs. JNC 7, published after our study, emphasizes that most patients will require more than one antihypertensive drug to achieve control.[9] In our study using data through the end of 2000, the majority of patients across all BP groups were already on at least two medications, with many also on a diuretic. Thus, we would not expect substantially different findings with more current data other than slightly lower rates of severe hypertension as noted above. Our measure of treatment adequacy does not account for medication dosage and thus may be overestimating treatment adequacy. It is possible that this may in part explain

the lack of statistical difference between the proportions controlled and uncontrolled in the severe hypertension group at study end. However, this still does not account for the fact that many subjects in the severe hypertension group were not even on adequate therapy based on number of medications.

Although BP control, especially in severe hypertension subjects, may be improved by treatment intensification, specialist referral and investigation for and management of secondary causes of hypertension, we were unable to examine reasons for lack of treatment intensification and lacked data on referrals to hypertension specialists (although we do know that less than 1.0% of these patients were seen in a hypertension clinic) or on investigations to exclude secondary hypertension. Other investigators have found clinical uncertainty about true BP values to be a prominent reason for lack of treatment intensification.[27] Whether this has a role in patients with such high BPs is unclear.

In conclusion, among veterans with hypertension, severe BP elevations are relatively common with many patients having persistent or recurrent elevations with inadequate follow-up and intensification of therapy. This suggests that clinical inertia is not just an issue among those with mildly elevated BP. Given the increased cardiovascular risk associated with this degree of BP elevation, there is a need for better understanding of how these patients are being managed. In addition, interventions are needed to overcome clinical inertia and improve providers' management of hypertension, especially regarding severe hypertension.

What is known about the topic

- Little data on prevalence and management of patients with severe hypertension.
- Prevalence estimates of 8–19% based on small sample sizes ($N = 252$–1413) in select populations (for example, patients referred to hypertension clinics).
- No data on patterns of severe hypertension and how long patients are exposed to such high BPs.

What the study adds

- Largest study to date examining prevalence and management of patients with severe hypertension ($N = 59\,207$, of whom 13 735 had at least one visit with severe hypertension).
- First to characterize the pattern of severe hypertension, that is, isolated spike vs persistent or recurrent.
- First to compare management of patients with severe hypertension to patients with other categories of hypertension.

Conflict of interest

This project was funded in part by the Department of Veterans Affairs Health Services Research and Development service grant number SDR 99-300-1 and by Bristol-Myers Squibb. Dr Berlowitz has served as a consultant for Bristol-Myers Squib. All aspects of design, conduct and analysis were under sole authority of the authors without any restriction on publication.

References

1 Kearney PM, Whelton M, Reynolds K, Whelton PK, He J. Worldwide prevalence of hypertension: a systematic review. *J Hypertens* 2004; **22**: 11–19.
2 Lewington S, Clarke R, Qizilbash N, Peto R, Collins R. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. *Lancet* 2002; **360**: 1903–1913.
3 SHEP Cooperative Research Group. Prevention of stroke by antihypertensive drug treatment in older persons with isolated systolic hypertension. Final results of the Systolic Hypertension in the Elderly Program (SHEP). *JAMA* 1991; **265**: 3255–3264.
4 Collins R, Peto R, MacMahon S, Hebert P, Fiebach NH, Eberlein KA *et al.* Blood pressure, stroke, and coronary heart disease. Part 2, short-term reductions in blood pressure: overview of randomised drug trials in their epidemiological context. *Lancet* 1990; **335**: 827–838.
5 Joint National Committee. The sixth report of the Joint National Committee on prevention, detection, evaluation, and treatment of high blood pressure. *Arch Intern Med* 1997; **157**: 2413–2446.
6 Whitworth JA. 2003 World Health Organization (WHO)/International Society of Hypertension (ISH) statement on management of hypertension. *J Hypertens* 2003; **21**: 1983–1992.
7 Mancia G, De Backer G, Dominiczak A, Cifkova R, Fagard R, Germano G *et al.* 2007 Guidelines for the Management of Arterial Hypertension: The Task Force for the Management of Arterial Hypertension of the European Society of Hypertension (ESH) and of the European Society of Cardiology (ESC). *J Hypertens* 2007; **25**: 1105–1187.
8 Guidelines Committee. 2003 European Society of Hypertension—European Society of Cardiology guidelines for the management of arterial hypertension. *J Hypertens* 2003; **21**: 1011–1053.
9 Chobanian AV, Bakris GL, Black HR, Cushman WC, Green LA, Izzo Jr JL *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* 2003; **289**: 2560–2572.
10 Salerno CM, Demopoulos L, Mukherjee R, Gradman AH. Combination angiotensin receptor blocker/hydrochlorothiazide as initial therapy in the treatment of patients with severe hypertension. *J Clin Hypertens (Greenwich)* 2004; **6**: 614–620.
11 Flack JM, Hamaty M. Difficult-to-treat hypertensive populations: focus on African-Americans and people with type 2 diabetes. *J Hypertens Suppl* 1999; **17**: S19–S24.
12 Lalljie GR, Lalljie SE. Characteristics and control of severe hypertension in a specialist, private practice in Jamaica. *West Indian Med J* 2005; **54**: 315–318.
13 Marquez-Contreras E, Coca A, de la Figuera von Wichmann M, Divison JA, Llisterri JL, Sobrino J *et al.*

Cardiovascular risk profile of uncontrolled hypertensive patients. The Control-Project study. *Med Clin (Barc)* 2007; **128**: 86–91.
14 Preston RA, Baltodano NM, Cienki J, Materson BJ. Clinical presentation and management of patients with uncontrolled, severe hypertension: results from a public teaching hospital. *J Hum Hypertens* 1999; **13**: 249–255.
15 Borzecki AM, Glickman ME, Kader B, Berlowitz DR. The effect of age on hypertension control and management. *Am J Hypertens* 2006; **19**: 520–527.
16 Borzecki AM, Wong AT, Hickey EC, Ash AS, Berlowitz DR. Can we use automated data to assess quality of hypertension care? *Am J Manag Care* 2004; **10**: 473–479.
17 VA Information Resource Center: Toolkit for new users of VA data Available at. http://www.virec.research.va.gov/Support/Training-NewUsersToolkit/Toolkit.htm. Accessed March 2008.
18 Borzecki AM, Wong AT, Hickey EC, Ash AS, Berlowitz DR. Identifying hypertension-related comorbidities from administrative data: what's the optimal approach? *Am J Med Qual* 2004; **19**: 201–206.
19 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): 613–619.
20 Berlowitz DR, Ash AS, Hickey EC, Friedman RH, Glickman M, Kader B *et al.* Inadequate management of blood pressure in a hypertensive population. *N Engl J Med* 1998; **339**(27): 1957–1963.
21 Borzecki AM, Wong AT, Hickey EC, Ash AS, Berlowitz DR. Hypertension control: how well are we doing? *Arch Intern Med* 2003; **163**: 2705–2711.
22 Hyman DJ, Pavlik VN. Characteristics of patients with uncontrolled hypertension in the US. *N Engl J Med* 2001; **345**: 479–486.
23 Hajjar I, Kotchen TA. Trends in prevalence, awareness, treatment, and control of hypertension in the US, 1988–2000. *JAMA* 2003; **290**: 199–206.
24 Cushman WC, Ford CE, Cutler JA, Margolis KL, Davis BR, Grimm RH *et al.* Success and predictors of blood pressure control in diverse North American settings: the Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT). *J Clin Hypertens* 2002; **4**: 393–405.
25 Berlowitz DR, Cushman WC, Glassman P. Hypertension in adults across age groups. *JAMA* 2005; **294**: 2970–2971;author reply 2971–2972.
26 Yu W, Ravelo A, Wagner TH *et al.* Prevalence and costs of chronic conditions in the VA Health Care System. *Med Care Res Rev* 2003; **60**(3 Suppl): 146S–167S.
27 Kerr EA, Zikmund-Fisher BJ, Klamerus ML, Subramanian U, Hogan MM, Hofer TP. The role of clinical uncertainty in treatment decisions for diabetic patients with uncontrolled blood pressure. *Ann Intern Med* 2008; **148**: 717–727.

CLINICAL RESEARCH STUDY

THE AMERICAN JOURNAL of MEDICINE®

Health Insurance and Cardiovascular Disease Risk Factors

Erica L. Brooks, MD,[a,b] Sarah Rosner Preis, ScD, MPH,[a,c] Shih-Jen Hwang, PhD,[a,c] Joanne M. Murabito, MD, ScM,[a,d] Emelia J. Benjamin, MD, ScM,[a,e] Margaret Kelly-Hayes, EdD, RN,[a,f] Paul Sorlie, PhD,[g] Daniel Levy, MD[a,c]

[a]The National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, Mass; [b]The Division of Cardiology, The Department of Medicine, Tufts Medical Center, Boston, Mass; [c]The Center for Population Studies of the National Heart, Lung, and Blood Institute, Bethesda, Md; [d]Section of General Internal Medicine and School of Medicine, Boston University, Boston, Mass; [e]School of Medicine and School of Public Health, Boston University, Boston, Mass; [f]Department of Neurology and School of Medicine, Boston University, Boston, Mass; [g]Division of Prevention and Population Sciences, National Heart, Lung, and Blood Institute, Bethesda, Md.

ABSTRACT

BACKGROUND: Compared with those with health insurance, the uninsured receive less care for chronic conditions, such as hypertension and diabetes, and experience higher mortality.

METHODS: We investigated the relations of health insurance status to the prevalence, treatment, and control of major cardiovascular disease risk factors—hypertension and elevated low-density lipoprotein (LDL) cholesterol—among Framingham Heart Study (FHS) participants in gender-specific, age-adjusted analyses. Participants who attended the seventh Offspring cohort examination cycle (1998-2001) or the first Third Generation cohort examination cycle (2002-2005) were studied.

RESULTS: Among 6098 participants, 3.8% were uninsured at the time of the FHS clinic examination and ages ranged from 19 to 64 years. The prevalence of hypertension and elevated LDL cholesterol was similar for the insured and uninsured; however, the proportion of those who obtained treatment and achieved control of these risk factors was lower among the uninsured. Uninsured men and women were less likely to be treated for hypertension with odds ratios for treatment of 0.19 (95% confidence interval [CI], 0.07-0.56) for men and 0.31 (95% CI, 0.12-0.79) for women. Among men, the uninsured were less likely to receive treatment or achieve control of elevated LDL cholesterol than the insured, with odds ratios of 0.12 (95% CI, 0.04-0.38) for treatment and 0.17 (95% CI, 0.05-0.56) for control.

CONCLUSION: The treatment and control of hypertension and hypercholesterolemia are lower among uninsured adults. Increasing the proportion of insured individuals may be a means to improve the treatment and control of cardiovascular disease risk factors and to reduce health disparities.

Published by Elsevier Inc. • The American Journal of Medicine (2010) 123, 741-747

KEYWORDS: Cardiovascular risk factors; Health disparities; Health insurance; Hypertension

The lack of health insurance is a large and growing problem in the United States. In 2008, 15.4% of the US population was uninsured, and the number of uninsured persons has increased over the past 2 decades.[1]

In the United States, adults without health insurance are less likely to receive screening for chronic medical conditions, such as hypertension and hypercholesterolemia.[2] Conversely, approximately 1 in 7 adults with diabetes or hypertension is without health insurance.[3] Compared with the insured near-elderly, the uninsured near-elderly have a higher mortality rate, which is mostly confined to those with diabetes, hypertension, and coronary heart disease.[4]

Of studies that have looked at the treatment and control of cardiovascular disease risk factors by health insurance status, the majority have relied on self-reported diagnoses,[5] only evaluated medication use without addressing the effectiveness of treatment,[6,7] or focused on hypertension alone.[7,8] The current investigation examined the prevalence, treatment, and control of hyper-

Funding: The Framingham Heart Study is funded through National Institutes of Health contract N01-HC-25195.

Conflict of Interest: None of the authors have any conflicts of interest associated with the work presented in this manuscript.

Authorship: All authors had access to the data and played a role in writing this manuscript.

Reprint requests should be addressed to Daniel Levy, MD, Framingham Heart Study, 73 Mount Wayte Ave, Framingham, MA 01702-5803.

E-mail address: levyd@nih.gov

0002-9343/$ -see front matter Published by Elsevier Inc.
doi:10.1016/j.amjmed.2010.02.013

tension and elevated low-density lipoprotein (LDL) cholesterol in Framingham Heart Study (FHS) participants according to health insurance status. The majority of the FHS participants reside in Massachusetts, which has a highly insured population compared with the US average and has legislated mandatory health insurance effective July 2007.[9] The average percentage of the population that was uninsured from 2004 to 2006 was 10.3% for Massachusetts compared with 15.3% nationally.[10] Demonstrating differences in the treatment of cardiovascular disease risk factors in insured versus uninsured individuals in an area with high health insurance rates would underscore the public health implications of expanded insurance coverage.

CLINICAL SIGNIFICANCE

- The prevalence of hypertension and elevated low-density lipoprotein cholesterol is similar among the insured and uninsured.
- The treatment and control of hypertension and hyperlipidemia are significantly lower for uninsured than insured men.
- Hypertension treatment is significantly lower in uninsured than insured women.
- The lack of control of cardiovascular risk factors may explain poorer health and increased mortality in the uninsured compared with the insured.

MATERIALS AND METHODS

Study Population

The FHS is an observational study that began in 1948 when an original study cohort was enrolled. The children and spouses of children of the Original cohort were enrolled in the Offspring cohort, beginning in 1971.[11] The Third Generation cohort, composed of children of Offspring cohort participants, was enrolled starting in 2002.[12] Data from the 3539 participants in the Offspring cohort's seventh examination cycle (1998-2001) and 4095 participants in the first examination cycle for the Third Generation cohort (2002-2005) form the basis of this investigation. Participants who were aged 65 years or more and therefore eligible for Medicare (n = 1238) or had missing data on cardiovascular disease risk factors (n = 245) were excluded. An additional 53 participants with missing data on insurance status were excluded, resulting in a final sample size of 6098 participants.

Exposure Variable

Insurance status was sought as part of the sociodemographic questionnaire that participants completed during the clinic visit. Offspring cohort participants were asked, "Do you have health insurance?" Third Generation participants were asked about various types of insurance, including Medicare, Medicaid, Veterans Administration or military insurance, private or health maintenance organization insurance, or no insurance. Participants were classified as having no insurance if they did not respond affirmatively to any of the insurance categories but did answer at least 1 item of the insurance question.

Measurements and Definitions of Outcome Variables

The main outcome measures were blood pressure and LDL cholesterol. Blood pressure was determined by the average of 2 measurements performed by the examining physician during the clinic visit. Participants with systolic blood pressure of 140 mm Hg or greater or diastolic blood pressure of 90 mm Hg or greater, or those taking medication for a hypertension indication were defined as having hypertension. Medication use was by self-report. All participants had fasting blood work performed at the clinic visit. LDL cholesterol was calculated using the Friedewald equation.[13] A diagnosis of elevated LDL cholesterol was defined by National Cholesterol Education Program Adult Treatment Panel III guidelines and was LDL cholesterol 160 mg/dL or more for those with no or 1 cardiovascular disease risk factor, 130 mg/dL or more for those with 2 or more risk factors, and 100 mg/dL or more for those with coronary heart disease, diabetes or a coronary heart disease risk equivalent,[14] or current use of a lipid-lowering agent.

Participants were diagnosed with coronary heart disease if they had prevalent myocardial infarction, coronary insufficiency, or angina as determined by physician investigator review of FHS clinic visit and hospital records. Similarly, participants were diagnosed as having cardiovascular disease if they had a diagnosis of coronary heart disease, stroke (ischemic stroke or intracerebral hemorrhage), heart failure, or intermittent claudication as determined by physician review. For those without prevalent cardiovascular disease, a Framingham risk score was calculated predicting the 10-year risk of coronary heart disease.[15] The presence of metabolic syndrome was defined as 3 or more of the following: waist circumference of 102 cm or more for men or 88 cm or more for women, triglycerides of 150 mg/dL or more, high-density lipoprotein cholesterol less than 40 mg/dL for men or less than 50 mg/dL for women, fasting blood glucose of 100 mg/dL or more or use of insulin or an oral hypoglycemic agent, and blood pressure of 130/85 mm Hg or more or antihypertensive medication use.[16]

Participants provided detailed information on medical checkups, hospitalizations, emergency department visits, smoking, and alcohol use during the physician-administered examination. Heavy alcohol use was defined as 7 or more drinks per week for women and 14 or more drinks per week for men.[17] The Center for Epidemiologic Studies Depression Scale score was used to determine depressive symptomatology with a score of 16 or greater indicating a high degree of symptoms.[18] Participants also answered questionnaires providing detailed health and sociodemographic information, such as self-reported education and income lev-

els (both carried over from previous examinations) and health status.

Treatment and Control of Outcome Variables

Hypertension treatment was defined as currently taking antihypertensive medication. Control of hypertension was defined as blood pressure less than 140/90 mm Hg. Treatment for elevated LDL cholesterol was defined as currently taking prescription lipid-lowering medication. Control of LDL cholesterol was defined by Adult Treatment Panel III guidelines and depended on the number of cardiovascular disease risk factors each participant had at the time of the examination with levels being less than those described above.[14] Treatment for all conditions was defined by the proportion of participants being treated for a condition among those with the condition. Similarly, control was defined as the proportion of participants with adequate control of a given risk factor among those defined as having that risk factor.

Statistical Analysis

All analyses were gender-specific because of the presence of spouse couples within the study sample that were unlikely to be independent with respect to insurance status. For all continuous variables, generalized linear models, adjusted for age and cohort (Offspring vs Third Generation), were used to compare differences in mean risk factor levels between the insured and uninsured. For dichotomous variables, a logistic regression model, adjusting for age and cohort (Offspring vs Third Generation), was constructed to compare the proportion of clinical and psychosocial factors among those with and without health insurance. Multinomial logistic regression was used for polytomous variables. To assess the relations between health insurance status and cardiovascular disease risk factor prevalence, treatment, and control, odds ratios and 95% confidence intervals were calculated using generalized estimating equations and logistic regression to adjust for age and relatedness between study participants, because the FHS has a family-based design. All statistical analyses were performed using SAS, v. 8.2 (SAS Institute Inc, Cary, NC). A 2-sided P value of less than .05 was considered statistically significant.

RESULTS

Baseline Clinical Characteristics

Women comprised 53.5% of the study sample. Five percent of men and 3% of women were uninsured; 59% of the uninsured were men. Mean systolic and diastolic blood pressures for the insured and uninsured were 122/78 and 121/77 mm Hg for men and 116/73 and 118/73 mm Hg for women, respectively (Table 1). For men, total cholesterol and LDL cholesterol values were significantly higher among the uninsured at 200 and 126 mg/dL compared with 194 and 121 mg/dL for the insured, respectively ($P = .01$ and $P = .03$). In women, no differences were seen in lipid concentrations according to health insurance status with mean LDL cholesterol values of 110 and 114 mg/dL for insured and uninsured women, respectively. The prevalence of preexisting coronary heart disease or cardiovascular disease, the average Framingham coronary heart disease risk

Table 1 Baseline Clinical Characteristics by Health Insurance Status for Men and Women

	Men			Women		
	Insured (N = 2697)	Uninsured (N = 136)	*P* value*	Insured (N = 3171)	Uninsured (N = 94)	*P* Value*
Age (y), mean ± SD	45 ± 11	42 ± 10	.01	45 ± 11	46 ± 11	.72
Systolic blood pressure (mm Hg), mean ± SD	122 ± 14	121 ± 15	.77	116 ± 16	118 ± 18	.3
Diastolic blood pressure (mm Hg), mean ± SD	78 ± 9	77 ± 10	.22	73 ± 9	73 ± 10	.97
Fasting blood glucose (mg/dL), mean ± SD	101 ± 20	101 ± 23	.26	94 ± 20	97 ± 24	.22
Total cholesterol (mg/dL), mean ± SD	194 ± 36	200 ± 40	.01	193 ± 36	195 ± 40	.67
LDL-C (mg/dL), mean ± SD	121 ± 31	126 ± 35	.03	110 ± 32	114 ± 35	.4
HDL-C (mg/dL), mean ± SD	46 ± 13	47 ± 12	.35	61 ± 16	59 ± 18	.15
Triglyceride level (mg/dL), mean ± SD	139 ± 103	141 ± 114	.51	106 ± 67	117 ± 68	.17
Framingham 10-y risk score,† mean ± SD	8.2 ± 7.0	7.0 ± 5.2	.73	3.5 ± 4.1	4.3 ± 4.6	.07
BMI (kg/m^2), mean ± SD	28.4 ± 4.7	28.1 ± 5.3	.98	26.5 ± 6.1	28.0 ± 7.0	.02
Obesity, n (%)	782 (29)	38 (28)	.83	721 (23)	32 (34)	.01
History of coronary heart disease, n (%)	88 (3)	4 (3)	.52	31 (1)	1 (1)	.98
History of cardiovascular disease, n (%)	115 (4)	5 (4)	.54	59 (2)	1 (1)	.49
Metabolic syndrome, n (%)	920 (34)	38 (28)	.59	626 (20)	24 (26)	.24
Current smoker, n (%)	443 (16)	62 (46)	<.001	492 (16)	31 (33)	<.001
Heavy alcohol use,‡ n (%)	404 (15)	28 (21)	.04	418 (13)	17 (18)	.23

BMI, Body mass index; LDL-C, low-density lipoprotein cholesterol; HDL-C, high-density lipoprotein cholesterol; SD, standard deviation.
*P values are adjusted for age (years) and cohort (Offspring vs Third Generation).
†Excludes those with known coronary heart disease.
‡Heavy alcohol use defined as ≥14 drinks/week in men and ≥7 drinks/week in women.

Table 2 Baseline Psychosocial and Socioeconomic Characteristics by Health Insurance Status for Men and Women

	Men			Women		
	Insured (N = 2697)	Uninsured (N = 136)	*P* Value*	Insured (N = 3171)	Uninsured (N = 94)	*P* Value*
CES-D score ± 16, n (%)	216 (8)	32 (24)	<.001	383 (12)	19 (20)	.02
Excellent or very good self-reported health status, n (%)	1725 (64)	74 (54)	.004	2129 (67)	44 (47)	<.001
Health care use, n (%)						
Emergency department visits	1288 (48)	82 (60)	.05	1332 (42)	44 (47)	.19
≥1 hospitalizations	1194 (44)	66 (49)	.74	1829 (58)	40 (43)	.006
Checkup by doctor in the last 5 y	2369 (88)	79 (58)	<.001	2976 (94)	71 (76)	<.001
>12 y of education, n (%)	1952 (72)	72 (53)	<.001	2336 (74)	58 (62)	.32
Family income ≥ $75,000, n (%)	921 (34)	16 (12)	<.001	985 (31)	3 (3)	<.001

CES-D, Center for Epidemiologic Studies Depression Scale.
**P* values are adjusted for age (years) and cohort (Offspring vs Third Generation).

score, and the prevalence of the metabolic syndrome did not vary by health insurance status for men or women. Uninsured women had a significantly higher mean BMI (28.0 vs 26.5; $P = .02$) and a higher prevalence of obesity (34% vs 23%; $P = .01$). In men, 46% of the uninsured smoked compared with 16% of insured men ($P < .001$); corresponding values for women were 33% and 16% ($P < .001$). Uninsured men were more likely to have heavy alcohol use than insured men (21% vs 15%; $P = .04$).

Baseline Psychosocial and Sociodemographic Characteristics

A comparison of categoric variables is presented in Table 2. There was a significant difference between the uninsured and insured in the percentage of persons defined as having a high level of depressive symptoms on the Center for Epidemiologic Studies Depression Scale, which was 24% versus 8% ($P < .001$) for men and 20% versus 12% ($P = .02$) for women. The rates of hospitalization were similar between uninsured and insured men; however, uninsured men were more likely to have been seen in the emergency department. Among women, the uninsured were less likely to have been hospitalized for any reason, including childbirth, 43% versus 58% ($P = .006$). The rates of routine physical examinations within the last 5 years were significantly lower among uninsured participants compared with insured participants in men (58% vs 88%; $P < .001$) and in women (76% vs 94%; $P < .001$). Only 53% of uninsured men obtained post high school education compared with 72% of insured men ($P < .001$). There were no significant differences in the percentage of women with post-high school education by health insurance status. The uninsured were more likely to have lower self-reported health status and had lower rates of family income $75,000 or greater than the insured.

Prevalence, Treatment, and Control of Cardiovascular Disease Risk Factors

Tables 3 (men) and 4 (women) report the gender-specific prevalence, treatment, and control of cardiovascular disease risk factors, adjusted for age and cohort. The mean systolic and diastolic blood pressures for men with a diagnosis of hypertension were 134/84 ± 16/10 mm Hg and 141/89 ±

Table 3 Cardiovascular Disease Risk Factor Prevalence, Treatment, and Control by Health Insurance Status for Men

	Insured, n (%)	Uninsured, n (%)	Odds Ratio* (95% CI)	*P* Value*
Hypertension				
Prevalence	758 (28)	25 (18)	0.72 (0.45-1.17)	.19
Treatment	449 (59)	4 (16)	0.19 (0.07-0.56)	.003
Control	332 (42)	2 (8)	0.17 (0.04-0.68)	.01
Elevated Low-Density Lipoprotein Cholesterol				
Prevalence	941 (35)	44 (32)	1.19 (0.79-1.80)	.41
Treatment	383 (41)	3 (7)	0.12 (0.04-0.38)	<.001
Control	287 (31)	3 (7)	0.17 (0.05-0.56)	.004

CI, Confidence interval.
*Odds ratios and *P* values are adjusted for age (years) and relatedness between family members.

Table 4 Cardiovascular Disease Risk Factor Prevalence, Treatment, and Control by Health Insurance Status for Women

	Insured, n (%)	Uninsured, n (%)	Odds Ratio* (95% CI)	*P* Value*
Hypertension				
Prevalence	580 (18)	21 (22)	1.19 (0.66-2.15)	.56
Treatment	397 (68)	8 (38)	0.31 (0.12-0.79)	.01
Control	287 (49)	7 (33)	0.57 (0.22-1.44)	.23
Elevated Low-Density Lipoprotein Cholesterol				
Prevalence	572 (18)	25 (27)	1.63 (0.99-2.67)	.05
Treatment	207 (36)	6 (24)	0.55 (0.22-1.40)	.21
Control	168 (29)	4 (16)	0.47 (0.16-1.36)	.16

CI, Confidence interval.
*Odds ratios and *P* values are adjusted for age (years) and relatedness between family members.

14/8 mm Hg for those with and without insurance, respectively. The corresponding values for women with hypertension were 136/81 ± 18/10 mm Hg and 137/83 ± 22/11 mm Hg. Although the prevalence of hypertension was similar in uninsured versus insured men and women, the proportion of those treated for hypertension was lower for those without health insurance. The odds ratios for hypertension treatment were 0.19 in men (uninsured vs insured; $P = .003$) and 0.31 in women ($P = .01$). The proportion of those with controlled hypertension also was significantly lower in uninsured men at 8% compared with 42% in insured men (odds ratio of control of 0.17; $P = .01$).

Mean LDL cholesterol values among men with a diagnosis of elevated LDL cholesterol were 137 ± 37 mg/dL and 154 ± 35 mg/dL for the insured and uninsured, respectively. Corresponding values in women were 142 ± 39 mg/dL and 149 ± 38 mg/dL. The pattern for high LDL cholesterol among men was similar to that of hypertension. The proportion of those treated for high LDL cholesterol was 41% in insured men but only 7% in uninsured men (odds ratio of treatment 0.12; $P < .001$). Control of LDL cholesterol was achieved in only 7% of uninsured men with elevated LDL cholesterol versus 31% in insured men (odds ratio of control 0.17; $P = .004$). The percentage of women with elevated LDL cholesterol was marginally higher in the uninsured than the insured, but no differences in treatment or control were observed.

DISCUSSION

The proportions of those with treated and controlled major cardiovascular disease risk factors were considerably lower in uninsured compared with insured individuals. This was most notable for treatment and control of hypertension and elevated LDL cholesterol in men and for hypertension treatment in women. Whereas the lower rate of hypertension control in the uninsured has been demonstrated,[5] the finding that the proportion of controlled hypercholesterolemia also is significantly lower in uninsured men than insured men is new.

Our investigation demonstrated lower proportions of treatment and control of blood pressure among uninsured hypertensive individuals. The only randomized insurance study in the United States, the RAND Health Insurance Experiment, demonstrated that hypertensive individuals randomized to free health care had better blood pressure control than those who were randomized to insurance plans that required cost sharing.[19] Although the RAND study did not randomize people to uninsured versus insured health insurance status, and instead compared plans with a range in the amount of cost sharing for participants, their findings were consistent with the current results. In 2 quasi-experimental studies of insurance status investigating instances when Medi-Cal and Veterans Administration health insurance benefits were terminated, it was found that hypertensive patients whose benefits were cut experienced subsequent increases in blood pressure compared with those whose coverage was maintained.[20,21] Taken as a whole, these prior studies and our findings suggest that the lack of health insurance does have direct adverse effects on blood pressure for those with hypertension.

The prevalence of elevated LDL cholesterol did not differ between uninsured and insured men and was marginally higher in uninsured versus insured women in this study. Uninsured men were significantly less likely to have their LDL cholesterol levels treated or controlled than insured men. To assess hypercholesterolemia, we focused our investigation on LDL cholesterol because it is the focus of clear diagnostic and treatment guidelines.[14] Of the previous studies that have examined prevalence of hypercholesterolemia by insurance status, all have used total serum cholesterol and not LDL cholesterol as in this evaluation.[5,6,22] The difference in methodology may partially explain the new findings in this investigation. Our study also used contemporary data, from a time period when lipid-lowering treatment recommendations were more aggressive than in previous years.

Depressive Symptoms and Sociodemographic Factors

A notable finding was that the uninsured had significantly higher levels of depressive symptoms than the insured. It has been shown that the depressed uninsured are less likely to receive treatment than the depressed insured[23] and that depression is more severe in the uninsured than the insured.[24] Differences in psychosocial factors such as depression may not reflect an association with lack of insurance, but rather may reflect other common factors such as differences in education and socioeconomic status. However, the significance of depression and other psychiatric illnesses as comorbidities in the uninsured deserves further investigation. Consistent with previous work, our study demonstrated that the uninsured have lower rates of routine medical checkups, lower income and self-reported health status, and higher rates of smoking.[2,4,5]

Potential Mechanisms for These Findings

There are many possible mechanisms for why those without health insurance would have lower proportions of treated and controlled hypertension and hyperlipidemia. Although the prevalence of hypertension was similar between those with and without insurance, this does not mean that the rates of diagnosis also were similar between groups. Given that those without health insurance are less likely to have routine medical examinations than the insured, hypertension and hyperlipidemia are likely underdiagnosed among the uninsured. Indeed, the uninsured are less likely to be aware of personal diagnoses of hypertension or hyperlipidemia than the insured.[5] Even if the uninsured were diagnosed with these conditions, treatment is dependent on access to continued medical care and control of risk factors is dependent on obtaining treatment. Thus, decreased rates of routine medical examinations among the uninsured could have det-

rimental effects on rates of diagnosis, treatment, and control of cardiovascular disease risk factors. In addition, hypertension and hypercholesterolemia are asymptomatic conditions, and the uninsured may be less inclined to seek screening or care for these conditions. The costs of physician visits, blood chemistry tests, and prescription medication likely explain much of the observed lower proportions of treated and controlled cardiovascular disease risk factors among the uninsured compared with the insured. However, many other measured and unmeasured factors, such as lack of adherence to medical regimens because of depressive symptoms, poor understanding of health conditions because of lack of a regular health care provider, and cultural attitudes pertaining to the health care system, affect the interplay between health insurance and cardiovascular disease risk factor treatment and control. These complex interactions are beyond the scope of the present investigation but merit further elucidation.

Our study has a number of strengths. Data from the FHS are rigorously collected, 99% of participants had fasting blood glucose chemistry tests, and physician investigators review all cardiovascular disease end points. Notably, this study used physician-measured blood pressure and obtained fasting laboratory values to define the main risk factors and their treatment and did not rely on self-reported diagnoses as in a preceding study.[5]

Because of the cross-sectional nature of this investigation, we were unable to demonstrate that the lack of health insurance has a causal relation to uncontrolled risk factors or increased cardiovascular disease risk. Other limitations of this study include low numbers of uninsured participants, which limited our power to demonstrate differences in some outcomes by health insurance status. The participants in the FHS are almost entirely white and reside mainly in Massachusetts. Although the lack of geographic and racial diversity of the study participants does limit the ability to generalize the results of the current investigation, it also eliminates race as a confounder. Also, FHS participants undergo periodic examinations that can result in referrals back to their personal physician. Thus, FHS participants may have more contact with the health care system, greater health literacy, and increased awareness of personal diagnoses of conditions, such as hypertension and hyperlipidemia, than the general population. Unfortunately, we were not able to assess whether participants were underinsured. Including the underinsured, such as those with catastrophic insurance coverage only, among those with health insurance in this study might alter the observed association between health insurance status and rates of treatment and control of hypertension and hyperlipidemia.

CONCLUSIONS

Our investigation emphasizes the relations between insurance status and cardiovascular disease risk factor prevalence, treatment, and control. Although we studied a highly insured population—less than 5% of FHS participants were uninsured compared with more than 15% in the general population[1]—multiple noteworthy differences were identified. More research is needed to determine whether the associations we observed are replicated in different samples with a greater proportion of uninsured participants. Improved management of these common and modifiable risk factors may be one way to reduce disparities in health care for the uninsured.

References

1. DeNavas-Walt C, Proctor BD, Smith JC. U.S. Census Bureau, Current Population Reports, P60-236, *Income, Poverty, and Health Insurance Coverage in the United States: 2008*. Washington, DC: U.S. Government Printing Office; 2009.
2. DeVoe JE, Fryer GE, Phillips R, Green L. Receipt of preventive care among adults: insurance status and usual source of care. *Am J Public Health.* 2003;93:786-791.
3. Davidoff A, Kenney G. *Uninsured Americans with Chronic Health Conditions: Key Findings from the National Health Interview Survey.* Washington, DC: Urban Institute; 2005.
4. McWilliams JM, Zaslavsky AM, Meara E, Ayanian JZ. Health insurance coverage and mortality among the near-elderly. *Health Aff (Millwood).* 2004;23:223-233.
5. Fowler-Brown A, Corbie-Smith G, Garrett J, Lurie N. Risk of cardiovascular events and death—does insurance matter? *J Gen Intern Med.* 2007;22:502-507.
6. Ayanian JZ, Zaslavsky AM, Weissman JS, et al. Undiagnosed hypertension and hypercholesterolemia among uninsured and insured adults in the Third National Health and Nutrition Examination Survey. *Am J Public Health.* 2003;93:2051-2054.
7. Moy E, Bartman BA, Weir MR. Access to hypertensive care. Effects of income, insurance, and source of care. *Arch Intern Med.* 1995;155:1497-1502.
8. Duru OK, Vargas RB, Kermah D, et al. Health insurance status and hypertension monitoring and control in the United States. *Am J Hypertens.* 2007;20:348-353.
9. The General Court of The Commonwealth of Massachusetts. An Act Providing Access to Affordable, Quality Accountable Health Care. Chapter 58 of the Acts of 2006. Available at: http://www.mass.gov/legis/laws/seslaw06/sl060058.htm. Accessed June 10, 2010.
10. DeNavas-Walt C, Proctor B, Smith J. US Census Bureau, Current Population Reports. *Income, Poverty, and Health Insurance Coverage in the United States: 2006.* Washington, DC: U.S. Government Printing Office; 2007:60-233.
11. Feinleib M, Kannel WB, Garrison RJ, et al. The Framingham Offspring Study. Design and preliminary data. *Prev Med.* 1975;4:518-525.
12. 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.
13. Friedewald WT, Levy RI, Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clin Chem.* 1972;18:499-502.
14. Executive Summary of The Third Report of The National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, And Treatment of High Blood Cholesterol In Adults (Adult Treatment Panel III). *JAMA.* 2001;285:2486-2497.
15. Wilson PWF, D'Agostino RB, Levy D, et al. Prediction of coronary heart disease using risk factor categories. *Circulation.* 1998;97:1837-1847.
16. Grundy SM, Cleeman JI, Daniels SR, et al. Diagnosis and management of the metabolic syndrome. An American Heart Association/National Heart, Lung, and Blood Institute Scientific Statement. Executive summary. *Cardiol Rev.* 2005;13:322-327.

17. Lichtenstein AH, Appel LJ, Brands M et al. Diet and lifestyle recommendations revision 2006: a scientific statement from the American Heart Association Nutrition Committee. *Circulation.* 2006;114:82-96.
18. Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. *J Appl Psych Measures.* 1977;1:385-401.
19. Keeler EB, Brook RH, Goldberg GA, et al. How free care reduced hypertension in the health insurance experiment. *JAMA.* 1985;254:1926-1931.
20. Lurie N, Ward NB, Shapiro MF, Brook RH. Termination from Medi-Cal—does it affect health? *N Engl J Med.* 1984;311:480-484.
21. Fihn SD, Wicher JB. Withdrawing routine outpatient medical services: effects on access and health. *J Gen Intern Med.* 1988;3:356-362.
22. Ford ES, Will JC, De Proost Ford MA, Mokdad AH. Health insurance status and cardiovascular disease risk factors among 50-64-year-old U.S. women: findings from the Third National Health and Nutrition Examination Survey. *J Womens Health.* 1998; 7:997-1006.
23. Wells KB, Sherbourne CD, Sturm R, et al. Alcohol, drug abuse, and mental health care for uninsured and insured adults. *Health Serv Res.* 2002;37:1055-1066.
24. Lesser IM, Leuchter AF, Trivedi MH, et al. Characteristics of insured and noninsured outpatients with depression in STAR(*)D. *Psychiatr Serv.* 2005;56:995-1004.

Comparing a Self-Administered Measure of Empathy with Observed Behavior Among Medical Students

Daniel C. R. Chen, MD, MSc[1], M. Elaine Pahilan, MPH[2], and Jay D. Orlander, MD, MPH[1,3]

[1]Section of General Internal Medicine, Evans Department of Medicine, Boston University School of Medicine, Boston, MA, USA; [2]Office of Medical Education, Boston University School of Medicine, Boston, MA, USA; [3]Medical Service, VA Boston Healthcare System, Boston, MA, USA.

PURPOSE: Studies show that measures of physician and medical students' empathy decline with clinical training. Presently, there are limited data relating self-reported measures to observed behavior. This study explores a self-reported measure and observed empathy in medical students.

METHOD: Students in the Class of 2009, at a university-based medical school, were surveyed at the end of their 2nd and 3rd year. Students completed the Jefferson Scale of Physician Empathy-Student Version (JSPE-S), a self-administered scale, and were evaluated for demonstrated empathic behavior during Objective Structured Clinical Examinations (OSCEs).

RESULTS: 97.6% and 98.1% of eligible students participated in their 2nd and 3rd year, respectively. The overall correlation between the JSPE-S and OSCE empathy scores was 0.22, $p<0.0001$. Students had higher self-reported JSPE-S scores in their 2nd year compared to their 3rd year (118.63 vs. 116.08, $p<0.0001$), but had lower observed empathy scores (3.96 vs. 4.15, $p<0.0001$).

CONCLUSIONS: Empathy measured by a self-administered scale decreased, whereas observed empathy increased among medical students with more medical training.

KEY WORDS: empathy; JSPE; OSCE.

J Gen Intern Med 25(3):200–2
DOI: 10.1007/s11606-009-1193-4

Empathy in the physician-patient relationship is the physician's ability to recognize a patient's perspectives and experiences, and convey such an understanding back to the patient.[1,2] This understanding allows patients to feel respected and validated,[3] promotes patient and physician satisfaction, and may improve patient outcomes.[1,4–7] Empathy is one of the Association of American Medical Colleges' (AAMC) goals for the development and education of altruistic and compassionate physicians.[8]Studies of medical students and residents suggest that empathy decreases with increased medical training.[9–11] These studies used self-administered measures of empathy with an uncertain correlation with actual empathic behavior. One study by Hojat et al. showed a modest positive correlation between their self-administered measure of empathy at the beginning of the 3rd year of medical school and program directors' assessment of these students' empathy during the end of internship 3 years later.[12] While it is known that physician self-assessment does not compare favorably to observed measures of competence,[13] psychometrically sound scales are thought to do better if they are validated against observable behaviors.

This study explores the relationship between a self-administered measure of empathy, the *Jefferson Scale of Physician Empathy-Student Version* (JSPE-S), and observed empathy, as evaluated by standardized patients during end of year Objective Structured Clinical Examinations (OSCEs).

METHODS

Study Participants

All students in the Class of 2009 at Boston University School of Medicine (BUSM) were eligible to participate. The BUSM curriculum is a traditional 4-year medical school with 2 years of preclinical study, with limited patient contact in the form of weekly physician shadowing for 10 weeks and a weekly patient interviewing and examination course for 6 weeks, followed by 2 years of clinical clerkships and electives.

Study Design

Class of 2009 students in their 2nd and 3rd year of medical school were asked to participate in a voluntary online survey measuring "student attitudes toward medicine" during March-April 2007 and March-April 2008. The survey was administered during their end of year Objective Structured Clinical Examinations (OSCEs).

The *Jefferson Scale of Physician Empathy-Student Version* (JSPE-S), is a self-administered 20-item instrument measuring components of empathy among health professionals in patient-care situations.[2] Respondents indicate their level of agreement on a seven-point Likert scale. Scores range from 20 to 140, with higher values indicating a higher degree of empathy.

Electronic supplementary material *The online version of this article (doi:10.1007/s11606-009-1193-4) contains supplementary material, which is available to authorized users.*

Received March 20, 2009
Revised September 11, 2009
Accepted November 5, 2009
Published online December 15, 2009

Participants also specified their gender, age, anticipated financial debt and likelihood of choosing various specialties.

Career specialty intentions were categorized into two groups, "People-oriented" specialties and "Technology-oriented" specialties.[11] Categorizations were based on categories determined in prior studies.[2,11]

Toward the end of the 2nd and 3rd year of medical school, students are required to take an Objective Structured Clinical Examinations (OSCEs) where they are assessed on clinical skills, including their doctor-patient interactions, by standardized patients. Second-year students complete three cases—two history taking and physical exam cases and one substance abuse case—while 3rd-year students complete six cases—specialty-oriented cases in Medicine, Pediatrics, Family Medicine, Psychiatry, Ob/Gyn, and Surgery (see online Appendix A for descriptions of student cases). Each student is rated on a five-point Likert scale for empathy for each case (see online Appendix B for descriptions of OSCE empathy question). All standardized patients were trained at the University of Massachusetts Medical School and were familiar with the examination material and empathy question, and many had several years experience in evaluating Boston University School of Medicine students.

Descriptive statistics and analyses of variance (ANOVA) were used to compare the JSPE-S scores among the different classes and categorized groups, while controlling for the effects of gender, age, anticipated financial indebtedness, and career preference. Post-hoc ANOVA pairwise comparisons were made using Tukey's HSD test. Correlations were made between JSPE-S and observed empathy scores. All computations were done with SAS statistical software version 9.1. This study was approved by the Boston University Medical Center Institutional Review Board.

RESULTS

The Class of 2009 had 167 students eligible in the 2nd year and 162 students eligible in the 3rd year to participate in the study; 97.6% and 98.1% of eligible students participated in their 2nd and 3rd years, respectively (Table 1).

The primary multivariate analysis of variance considered and adjusted for five factors: class, gender, anticipated financial debt, career preference, and age. The overall correlation between JSPE-S and OSCE observed empathy scores was 0.22 ($p<0.0001$).

Second-year students had higher JSPE-S scores compared to 3rd-year students (118.63 vs. 116.08, $p<0.0001$), but the average observed empathy score for 2nd-year students was lower than the observed empathy score for 3rd-year students (3.96 vs 4.15, $p<0.0001$) (Table 2).

Table 2. Results by OSCE[a]

	Second-year assessment (N=163)	Third-year assessment (N=159)	P-value
JSPE-S average[b]	118.63	116.08	<0.0001
OSCE empathy average per case[b]	3.96	4.15	<0.0001

[a]*Adjusted for gender, age, anticipated financial indebtedness, and career preference*
[b]*Possible JSPE-S scores range from 20 to 140*
[c]*Possible OSCE scores range from 1 to 5*

DISCUSSION

Self-administered tools are the most common method used to assess various educational interventions aimed at improving student empathy.[14–20] Nevertheless, the validity of such self-assessments is unknown.[13] Our study found a trend towards a decline in measured empathy with increased clinical training with a self-administered instrument, but an improvement in observed empathy among these more clinically experienced students. Our finding of empathy decline by the JSPE-S is similar to prior studies using this instrument to assess change from the pre-clinical and clinical years.[10,11] So why the discrepancy with simulated patient ratings?

We can suggest several explanations. The differences noted between the 2nd- and 3rd-year medical school classes on their self-assessment of empathy could be attributed to different training experiences and how these experiences shape student perception of illness, compassion, and empathy. The JSPE-S is designed to assess the empathy of health-care providers in patient-provider situations. In the first 2 years of medical school, students interact with patients mostly when shadowing practicing physicians and participating in clinical skills training courses. Third-year students continually interact with patients and may often share the experiences of patients and families coping with serious and sometimes fatal illness. Therefore, it is possible that with their limited clinical exposure 2nd-year students interpret the subjective anchors of JSPE-S questions differently from 3rd-year students who have a broader and more intense clinical experience. The student, in effect, has recalibrated his emotional understanding of illness through clinical experience and hence altered his score on paper without having impacted his nature.

An alternative explanation could be that student acculturation to critical illness and a true emotional recalibration within the student. Such a change could be protective in the professional development of a physician. Medical illness and patient suffering are real, intense, and frequently sad. It may be necessary for physicians to undergo a professional acculturation that is being captured by the self-assessment tool and

Table 1. Demographics and Characteristics of the Students in the Medical School Class of 2009

	Second year	Third year
Number of students eligible to participate	167	162
Number of completed surveys	163	159
Percentage of surveyed students who were female	54.0% (N=88)	56.6% (N=90)
Percentage of surveyed students preferring "people-oriented" specialties	51.0% (N=78)	58.3% (N=88)
Percentage of eligible students surveyed	97.6%	98.1%

that, to date, has been interpreted as a measured decline in empathy associated with early clinical training. Students in this case do become emotionally hardened and feel less empathy. The working hypothesis has been that clinical training lacks an element of humanity and that a "hidden curriculum" exposed by jaded, experienced practitioners undermines the idealism, humanism, and empathy young clinicians bring with them to clinical medicine.[11,14] This last argument has been the motivation for curricular innovations designed to enhance and support the maintenance of empathy. It would be difficult to discern the difference between these two hypotheses, but each would predict potentially lower scores on self-reported empathy as one traverses the early stages of clinical training.

So why did we observe more empathic behavior in more advanced students? The rating of empathy during an OSCE broadly includes verbal and non-verbal communication, as well as physical behavior, but cannot assess the internal emotion or motivation of the student. Third-year students had more opportunities to practice, observe, and get feedback on their empathic behaviors, independent of their internal emotional connection. From the perspective of clinical proficiency, one would anticipate that more experienced students would better demonstrate clinical behaviors, including empathic behaviors.

Thus, the apparent independence of the self-assessment measure and observer ratings suggests that the use of the JSPE as self-assessment tool may not sufficiently predict empathic behavior. Further confirmation of our findings is needed as this has implications for curriculum evaluation since a self-administered tool is easier and cheaper to use. These findings raise more challenging questions for educators: What are the important aspects of physician empathy to measure? Is assessment of empathic behavior adequate even if internal emotions are discrepant? Is the correct attitude acceptable if we cannot relate it to competent behavior?

There are several limitations of our study. The small number of OSCE cases, especially in the 2nd year, raises issues of score reliability and, though OSCE score differences were present, it is unclear how clinically significant a 0.2 point difference is at the physician-patient level. Since we had only a single empathy question for each case, we are unable to determine which observable behaviors were driving the rating. Our study is limited to one medical school, but we feel that our results are applicable to all schools with a similar traditional structure. Future studies should examine the subtleties of the physician-patient interaction by discriminating those elements that comprise empathic behavior, such as tone of voice, empathic language, and non-verbal communication.

The patients' need for an empathic physician will always be essential. Efforts to improve the empathic behaviors of trainees are important. More work is required so that curricular enhancements designed achieve these goals can be properly evaluated.

Acknowledgment: *Permission to use the JSPE-S was obtained from the Jefferson Medical College Center for Research in Medical Education and Health Care.*

Funding Sources: *None of the authors received any funding support for the study.*

Conflict of Interest Statement: *None of the authors reported any financial conflicts of interest or have any affiliations relevant to the manuscript subject.*

Corresponding Author: *Daniel C. R. Chen, MD, MSc, Section of General Internal Medicine, Evans Department of Medicine, Boston University School of Medicine, 801 Massachusetts Avenue, 2nd Floor, Boston, MA 02118, USA (e-mail: daniel.chen@bmc.org).*

REFERENCES

1. **Coulehan JL, Platt FW, Egener B, et al.** "Let me see if I have this right…": words that help build empathy. Ann Intern Med. 2001;135:221–7.
2. **Hojat M, Mangione S, Nasca TJ, et al.** The Jefferson Scale of Physician Empathy: development and preliminary psychometric data. Educ Psychol Meas. 2001;61:349–65.
3. **Beckman HB, Markakis KM, Suchman AL, Frankel RM.** The Doctor-Patient relationship and malpractice: lessons from Plaintiff depositions. Arch Intern Med. 1994;154:1365–70.
4. **Bikker AP, Mercer SW, Reilly D.** A pilot prospective study on the consultation and relational empathy, patient enablement, and health changes over 12 months in patients going to the Glasgow Homoeopathic Hospital. J Altern Complement Med. 2005;11:591–600.
5. **MacPherson H, Mercer SW, Scullion T, Thomas KJ.** Empathy, enablement, and outcome: an exploratory study on acupuncture patients' perceptions. J Altern Complement Med. 2003;9:869–76.
6. **Vermeire E, Hearnshaw H, Van Royen P, Denekens J.** Patient adherence to treatment: three decades of research. A comprehensive review. J Clin Pharm Ther. 2001;26:331–42.
7. **Suchman AL, Roter D, Green M, Lipkin M.** Physician satisfaction with primary care office visits. Collaborative study group of the American Academy on physician and patient. Med Care. 1993;31:1083–92.
8. Association of American Medical Colleges: Medical School Objectives Project. http://www.aamc.org/meded/msop/start.htm. Accessed November 4, 2008.
9. **Bellini LM, Shea JA.** Mood change and empathy decline persist during three years of internal medicine training. Acad Med. 2005;80:164–7.
10. **Hojat M, Mangione S, Nasca TJ, et al.** An empirical study of decline in empathy in medical school. Med Educ. 2004;38:934–41.
11. **Chen D, Lew R, Hershman W, Orlander J.** A cross-sectional measurement of medical student empathy. J Gen Intern Med. 2007;22:1434–38.
12. **Hojat M, Mangione S, Nasca TJ, Gonnella JS, Magee M.** Empathy scores in medical school and ratings of empathic behavior in residency training 3 years later. J Soc Psychol. 2005;145:663–72.
13. **Davis DA, Mazmanian PE, Fordis, et al.** Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. JAMA. 2006;296:1094-102.
14. **Kramer D, Ber R, Moore M.** Increasing empathy among medical students. Med Educ. 1989;23:168–73.
15. **Anderson R, Schiedermayer D.** The art of medicine through the humanities: an overview of a one-month humanities elective for fourth year students. Med Educ. 2003;37:560–2.
16. **DasGupta S, Charon R.** Personal illness narratives: using reflective writing to teach empathy. Acad Med. 2004;79:351–6.
17. **Hatem D, Ferrara E.** Becoming a doctor: fostering humane caregivers through creative writing. Patient Educ Couns. 2001;45:13–22.
18. **Feighny KM, Monaco M, Arnold L.** Empathy training to improve physician-patient communciation skills. Acad Med. 1995;70:435–6.
19. **Stepien KA, Baernstein A.** Educating for empathy: a review. J Gen Intern Med. 2006;21:524–30.
20. **Winefield HR, Chur-Hansen A.** Evaluating the outcome of communication skill teaching for entry-level medical students: does knowledge of empathy increase? Med Educ. 2000;34:90–4.

JOURNAL OF WOMEN'S HEALTH
Volume 19, Number 4, 2010

DOI: 10.1089/jwh.2009.1591

Hierarchy as a Barrier to Advancement for Women in Academic Medicine

Peter Conrad, Ph.D.,[1] Phyllis Carr, M.D.,[2] Sharon Knight, Ph.D., R.N.,[3] Megan R. Renfrew, M.A.,[4] Mary B. Dunn, Ph.D.,[5] and Linda Pololi, MBBS, MRCP[6]

Abstract

Background: Research on barriers to professional advancement for women in academic medicine has not adequately considered the role of environmental factors and how the structure of organizations affects professional advancement and work experiences. This article examines the impact of the hierarchy, including both the organization's hierarchical structure and professionals' perceptions of this structure, in medical school organization on faculty members' experience and advancement in academic medicine.
Methods: As part of an inductive qualitative study of faculty in five disparate U.S. medical schools, we interviewed 96 medical faculty at different career stages and in diverse specialties, using in-depth semistructured interviews, about their perceptions about and experiences in academic medicine. Data were coded and analysis was conducted in the grounded theory tradition.
Results: Our respondents saw the hierarchy of chairs, based on the indeterminate tenure of department chairs, as a central characteristic of the structure of academic medicine. Many faculty saw this hierarchy as affecting inclusion, reducing transparency in decision making, and impeding advancement. Indeterminate chair terms lessen turnover and may create a bottleneck for advancement. Both men and women faculty perceived this hierarchy, but women saw it as more consequential.
Conclusions: The hierarchical structure of academic medicine has a significant impact on faculty work experiences, including advancement, especially for women. We suggest that medical schools consider alternative models of leadership and managerial styles, including fixed terms for chairs with a greater emphasis on inclusion. This is a structural reform that could increase opportunities for advancement especially for women in academic medicine.

Introduction

The advancement of women in academic medicine has lagged relative to their increased presence in medicine. The percentage of women in medical school has increased steadily over the past 30 years,[1] with the result that women constitute approximately half of medical school graduates,[2] yet the gender distribution of faculty in leadership positions in academic medicine remains primarily unchanged. For example, in terms of academic rank distribution by gender, among clinical scientists, 29% of male faculty compared with 14% of female faculty achieve full professorship positions[3] (only 17% of full professorships are held by women).[4] Women are somewhat more represented at associate professor (15% men vs. 6% women) and assistant professor levels (24% men vs. 17% women).[5] This was virtually unchanged from 2003 to 2008. In 2007, the average department chair's per medical schools were 21 male to 3 female chairs, a 7-fold difference.[6] As of 2008, 14 women were deans or interim deans (11%) of the current 129 medical schools[7]; interim deans are not guaranteed to assume deanship.

In addition to inequalities in rank and leadership, women are also paid less than men at the same rank[8–10] and move through the ranks of leadership more slowly when they do

[1]Department of Sociology, Brandeis University, Waltham, Massachusetts.
[2]Department of Medicine, Boston University School of Medicine, Boston, Massachusetts.
[3]Department of Health Education and Promotion, East Carolina University, College of Health and Human Performance, Greenville, North Carolina.
[4]Disparities Solutions Center, Institute for Health Policy, Massachusetts General Hospital, Boston, Massachusetts.
[5]Management Department, McCombs School of Business, The University of Texas at Austin, Texas.
[6]Women's Studies Research Center, Brandeis University, Waltham, Massachusetts.

advance.[9,11–13] According to a recent study, women earned 11% less than men, after adjusting for rank, track, degree, specialty, years in rank, and administrative positions.[14] Several studies have shown that some women in academic medicine also report experiencing gender-based discrimination.[14–16] Gender-based inequalities in pay and difficulties in professional advancement are found in other professions as well.[17,18]

Sociologists have made major progress establishing that race and gender matter at work; however, there has been less success in explaining why workers' sex and race affect their employment outcomes.[19] Much research on women's advancement (or lack of it) in professions has focused on cultural and gendered values. Other studies have examined how the structure of the organization relates to professional advancement and work experiences. One area of organizational interest is the impact of hierarchy on advancement. Kanter, in "The Impact of Hierarchical Structures on the Work Behavior of Women and Men," emphasized the importance of understanding how structures of power and hierarchical arrangements relate to inequity in the workplace.[20] Kanter's study shifts perceived differences in men's and women's work orientations from individual-level factors connected with the culture and socialization of women (family and work roles) to the role of organizational structures (e.g., organization hierarchy) in shaping apparent sex differences in the workplace (e.g., low aspirations as a result of opportunity structure).

Sex composition is one aspect of social structure that can affect social inequality.[21–24] Kanter's theory of tokenism suggests that the relative number of women and underrepresented minorities (URMs) can affect employment outcomes.[21,22] The proportion of women in leadership positions can have an effect on women's hiring and promotion when they are present in large enough numbers to form coalitions and affect personnel decisions.[25] Institutional practices, such as leadership terms and policies for transparency in decision making, can also affect social inequality in the workplace. Policies that limit the effects of decision makers' biases on employment outcomes can limit discrimination based on gender and race. The potential for bias is greatest when decision makers have full discretion over their selections.[19,26,27] Because institutional practices can have serious consequences for hiring and promoting women and URMs, they necessitate further study.

The importance of hierarchy and institutional practices in understanding gender-based work inequalities in medicine is underexplored,[20] with a focus on coping strategies[16] and the impact of gender-based unconscious biases on women's advancement into leadership positions.[28] Few articles have specifically examined how the hierarchy of medicine and specific institutional practices impact women's professional advancement and work experiences.[16,28]

As part of a larger interview study on women and advancement in academic medicine, we have discovered several cultural factors that appear to affect potentials for women faculty advancement.[29,30] This article aims to fill a gap in existing literature on barriers to advancement of women in academic medicine by linking discussions of inequality to the institutional structural barriers. Specifically, we explore the perceived impact of one structural factor, hierarchy in medical school organization, on women faculty's experience and advancement in academic medicine.

Materials and Methods

The data were collected as part of a study on the advancement of women and URM faculty in academic medicine, C-Change (The National Initiative on Gender, Culture and Leadership in Medicine). Five medical schools were selected representing diverse characteristics of U.S. medical schools. The schools were drawn from different regions, including two public and three private schools. The demographics of women and URM faculty were nearly identical to national statistics. The study was IRB approved.

Participant criteria

Stratified purposeful and chain referral strategies were used to identify and select medical faculty from the five C-Change medical schools according to school, gender, race/ethnicity, department/discipline, and career stage. The principal investigator (L.P.) obtained a confidential list of faculty from each school and selected participants based on these criteria to produce a stratified sample based on demographics, positions, and career stages. Participants included medical and surgical subspecialist, generalist, and research scientist faculty, with 84% having an M.D. terminal degree and 16% a Ph.D. A total of 96 faculty were interviewed, divided into four career stages: (1) early career (2–5 years as faculty), (2) plateaued (those who had not advanced as expected in rank and responsibility and had been faculty members for $\geq$10 years, (3) leadership (senior) faculty, including deans, department chairs, and center directors, and (4) left (former faculty who had left academic medicine). Interviewees were divided almost equally among the four groups and the five schools, but we interviewed fewer participants in the early career stage because we reached data saturation in this category early in the study.

Sample selection

A total of 175 faculty were invited to participate, 8 refused primarily because of time constraints, 54 never responded, and 12 others responded but were unable to be scheduled. Male plateau faculty were more difficult to identify than similar stage female faculty. Women (55%) and URM faculty were oversampled (17% African American, 4% Hispanic/Latino, and 79% Caucasian/white), as were generalist physicians (20%). Details on the breakdowns of gender, race, and stage of sample are available elsewhere.[29]

Data collection and analysis

Four of the authors (P.C., P.C., L.P., S.K.) conducted in-depth, open-ended interviews with the selected respondents. All were experienced interviewers and used the same research protocol when interviewing respondents; 15% of the interviews were conducted in person, the rest by telephone. Interviews, typically 1 hour in length, were audiorecorded and transcribed verbatim. Interviewers used an interview guide with 20 open-ended questions and dozens of probes to supplement the major questions, including items related to choice of medicine as a career, faculty aspirations, energizing aspects of their careers, advancement and advancement barriers, collaboration, leadership, power, values alignment, and work-family integration. The interview guide included no specific questions on hierarchy, but respondents discussed

hierarchy-related issues when answering questions about belonging, frustrations, decision making, power and leadership, and aspirations. Hierarchy-related issues emerged as a major concern through the coding process of the data.

The data were coded, and all names and identifying information were removed. Multiple coders compared, verified, and refined coding categories. Data were analyzed by repeated readings of interview transcripts with an analytic focus on understanding and interpreting meaning. Over 4000 pages of transcribed narrative were stored, coded, and sorted using Atlas.ti software. Analysis involved data condensation to identify patterns and themes emergent from the coded data. The analysis was inductive and data driven, in line with the grounded theory tradition.[31,32] To verify data patterns and conclusions, we continuously reviewed transcripts and discussed findings among co-authors to achieve consensus. In this article, participants are identified by gender, degree, and faculty category.

Results

The hierarchal organization of the medical school emerged as an issue of concern in our interviews. Many of our respondents view the medical school as a hierarchical institution that strongly impacts their experience in academic medicine. Sometimes, the faculty members we interviewed felt they were treated more like underling employees than professionals or colleagues. Many thought they were informed about decisions that affected their work lives rather than being active participants in the decision-making process; they did not believe they had adequate input in some decisions that were directly consequential to their work. Respondents often noted that the medical school was a very bureaucratic organization and apparently becoming more so. Sometimes, the hierarchy and organization seemed convoluted when faculty had to wend their way to get some information or a decision. These bureaucratic layers can lead to frustration, resentment, or even apathy. As an early career female Ph.D. faculty noted, "It's such a heavily tiered administrative monster, the medical school tiers." A male former academic physician described the hierarchy well:

> Well, the leadership—it's a very small academic department—there's a hierarchy of a chairman of the department, supervising division chiefs, and to become a division chief, seniority is very important, but also the amount of grant money you bring to the institution. It's almost ironic because frequently I see the people who are the best researchers are often named division chiefs, and these are not always the people who have the best managerial skills, but they have brought in the most research dollars to the academic institution, and it's unfortunate that frequently they have to give up some of their research in order to take on the administrative duties of division chief.

Length of tenure of chairs

One issue that came up often was the power and extended length of service of the chairs. To respondents, department chairs seemed to be appointed for indefinite terms and serve at the dean's pleasure or until they chose to step down (or up). As one male faculty noted:

> You are chair for life. I mean, you don't serve at the pleasure of the clock; you serve at the pleasure of the dean. And if it pleasures him for you to remain as chair for the remainder of your mortal days, you will remain as chair for the remainder of your mortal days. . . . I could resign. [male, Ph.D., leader]

The extended duration of chair appointments seems to be a real issue in the accumulation of power and authority in the hierarchy. One plauteaued female physician faculty noted that her department had had only three chairs since it was founded in the 1960s! Another pointed out that "removing a chair is a rare thing." [male, Ph.D., leader]

Styles of chairs

Numerous faculty said that virtually all important decisions are made by the department chair. Faculty often feel excluded at this level of decision making, even about decisions that affect their work lives (e.g., when and where to move offices, strictures related to changes in clinical responsibilities). Although there does seem to be more participation in decision making at the local or division level, the department seems very hierarchical and even more so at the upper administrative level. For example, a male physician stated, ". . . it's very hierarchical [so] those at the lower levels have minimum input I think by and large. Certainly not into major strategic decisions at a departmental level. It's all held at a very high level. . . ." [male, physician, leader].

Chairs varied greatly in their style of inclusiveness. Some chairs are rather authoritarian, even dictatorial, in their style of running a department, allowing no opposing viewpoints. One plateaued female physician said: ". . . we work in a department where if my department chair got word of what I was saying to you, it would threaten my position." At least one person (male, Ph.D. left) called this "a feudal system where the lords reign." The chair seems to set the tone for leadership in the organization, and his or her particular management style affects the experience of being in the department. For example, one physician described how different managerial styles set the tone of expected interactions and decision-making processes: "One is I'm the boss, talk to me; the other is I'm the boss, don't hesitate to talk to somebody who's keeping me from ever having to talk to you" (male, M.D., leader). Some chairs do adopt a more collegial, or at least inclusive, style of management. The "I'm the boss, talk to me" does not necessarily mean that decisions are made democratically (male, physician, leader) but sets more of a collaborative tone that is appreciated by department members. So participation often results from the individual style of the chair.

There are consequences to the hierarchy. Some faculty members thought that upper administration "doesn't have a clue" of what is happening in their division. As a female physician noted:

> I think the upper administration does not appear to be aware of the problems we have, which I think is very strange because I think at one point they had to be where we are now. . . . It's like parents don't get teenagers anymore and they were once teenagers [female, physician, early career]

How people move in and out of leadership positions (including advancement and tenure) is often described as a mystery, something "done behind closed doors" (male, Ph.D., left) or in a "black box" (male, physician, leader).

Some faculty see the pitfalls of the hierarchy and believe a more collegial organization might be more productive. As one woman (female, former faculty, Ph.D.) noted:

> So if [the organization] were actually function driven . . . or if our purpose is to produce really excellent physicians and an inspired group of people who want to do research and biomedicine, what kind of structure would that . . . it would look a whole lot different . . . a whole lot less hierarchical, a whole lot more collaborative and it would be a little more welcoming to women and minorities.

In sum, hierarchy, length of chairship, styles of leadership, and probably communication all affect leadership in medical schools in complex but not incomprehensible ways.

Hierarchy and transparency

One of the major consequences of the hierarchal structure is its impact on decision making. For numerous decisions, faculty experienced a lack of transparency. Inclusion varied by the level of decision making. Overall, faculty feel more involved in decisions on a local level (e.g., the clinic or the division) or in the decision-making process closer to their own work. The amount of involvement in decision making was almost a continuum, from medical school to department to division to clinical or teaching situation, and as several respondents noted (e.g., male, physician, plauteaued), inclusion and usually transparency depended on what kinds of decisions were being made. Numerous respondents recognized that many institutional decisions were complex and difficult to judge "because you're not privy to the information" (male, physician, left). Although the same individual noted, as he moved up in the organization, that "it's unbelievable what's available to me, which I was never privy to . . ." This is mediated in part, however, by the style of the chair, chief, or director. If the chair is one who includes people's views and opinions in decisions, faculty are more likely to believe they are part of the decision-making process. If the chair is authoritarian or chooses only a small group to consult, others will feel excluded. For example: "[Decisions are made] behind the scenes. A few key people deciding how to make something happen . . ." (male, physician, leader) or "decisions were made by a group of privileged individuals behind closed doors" (female, physician, plauteaued). A well-placed male physician in a leadership role commented how his chair set the tone: "There's not a lot of democratic decision making going on in my department." On the other hand, sometimes a new chair can bring in a new style, as an early career female Ph.D. noted about the consensus oriented style of the current chair. Similarly, a former faculty male Ph.D. described how transparency has gotten better with the current head. He noted, "If the head happens to be a good communicator and if the head chooses to seek input, you can feel a little like you've got some say in the matter. Otherwise, it's essentially decisions are made behind closed doors and you're informed." A plauteaued female physician pointed out how a change in chair and style transformed a close-knit department to a place where decisions are no longer shared and are made behind closed doors. It felt to her like a "loss of family," but as a former faculty male Ph.D., commented, sometimes these decisions turn out well and are actually benevolent. People still didn't see the process as adequately transparent, however.

Lack of inclusion and transparency occurred on higher levels as well. For the most part, people did not feel much involved about decisions on the university level. A female faculty recalled:

> It used to be that . . . decisions were not talked about openly and that decisions were made by some group of privileged individuals behind closed doors . . . and you always got the feeling that you were getting [only] part of the story. [female, physician, plauteaued]

Another male leader physician noted that the chancellor ran an "efficient, crisp, and clean and military-like organization, but the faculty felt excluded." Another respondent observed, "In our faculty meetings, it's not a discussion about (an issue) . . . It's usually a reporting out about decisions that have already been made" (female, physician, plauteaued).

Sometimes respondents thought that a small group was making the decisions. Despite feeling involved on the local (division) level, more than one faculty expressed not feeling involved in the department because it was "a very authoritarianly run department" (female, physician, plauteaued). Another woman mentioned that at the institutional level, decisions often felt arbitrary, but she still felt part of the decision making "at the level of the clinic, not at the level of finances. . . ." (female, physician, early career). As an early woman faculty member stated, "There are too many things that are unspoken. There are too many things that are not transparent."

It seems that people are most resentful when decisions are made without them that affect their lifestyle (e.g., call schedule, patient responsibilities) or immediate work environment. This is likely significant because people believe it is important to maintain some measure of autonomy and control in their everyday work. As a female early career physician described, "I don't have any authority about some other things, like, right now we're in a position where we need to hire." This is clearly a major issue for some people. One faculty noted that all the medical director positions were eliminated without any consultation, and this led to a problem about who would pick up the medical directors' salaries. She said, "I felt betrayed" (female, physician, plauteaued). Another faculty noted that after a doctoral program was eliminated, the school just said, "Fine, we won't replace you, we'll just give those students to [name]" (female, PhD, leadership). She felt they dumped the students on her, knowing she was the kind of person who would not let the students flounder.

As faculty move up the ranks in the hierarchy (and a few do), there is some sense they are more involved in decision making. As one woman pointed out, "In my administrative position now, I feel that I am actually consulted more. . . ." (female, Ph.D., leader). A female administrator noted, "Because I'm department chairman, I'm part of the council of clinical chiefs . . . so I feel that my voice is heard" (female, physician, leader). A male faulty member reflected that as he moved up the ranks, he felt more included, to the point where he now felt the decision making was "collaborative and inclusive" (male, physician, leader).

Hierarchy and gender

Both men and women generally described hierarchy, especially the structure of chairs, as a strong feature of academic medicine, but men seemed to be more tolerant of the structure (e.g., "it's just a different management style") and seemed less bothered by it than women. Men typically described the hierarchy in a very matter-of-fact fashion. For example, a male faculty member described the structure of academic medicine

and the lack of democratic decision making but did not suggest it negatively impacts him:

> It's very hierarchical and those at the lower levels have minimal input, I think, by and large. Certainly not into major strategic decisions at a department level. It's all held at a very high level, really, and partially, that reflects the way the chairman operates, and he doesn't really broker a lot of discussion about decisions. There's not a lot of democratic decision making going on in my department. [male, physician, leader]

However, another male physician, despite his leadership status, described feeling "excluded" yet noted some ironic benefits of being able to work around this challenge: "The bad news is that I feel excluded. The good news is he doesn't speak to me so he doesn't tell me what to do. So I go about my business in my area of responsibility for the most part" (male, physician, leader).

When male faculty talked about hierarchy and top down decisions, they talked more about poor communications with underlings as a problem. As one man observed: "I think you have a few people who are administrators . . . and I think one of the most simple things they forget is just communication, good communication with people who really do the work and pay the bills. . . ." (male, physician, early career).

Women, on the other hand, perceived the hierarchical structure to be more burdensome. Female faculty described very specific and detailed accounts of how system-level hierarchical processes (e.g., appointment processes) and gendered-cultural values negatively impacted their career progress and advancement. For example, one female faculty member commented on how she believed a normal search process was altered to specifically detour her from being appointed.

> Historically, what happened at the institution is that there's a division that needs a director. If there are senior productive people . . . it is suggested that they apply. . . . When it was time for the director to step down they said . . . we're not appointing anybody, we're going to conduct a national search, which of course they had never done before . . . it had been unprecedented. . . . I think it was people being a bit uncomfortable with me being appointed division director.

Several women presented stories about how they felt marginalized and not taken seriously (with some exceptions, of course). For example, female faculty reported feeling as if they were treated like teenagers (female, physician, early career) or singled out as "disruptive" to the department when they spoke up (female, physician, left). Women, unlike men, discussed whether or not their voice was heard within the medical department. They state several reasons for this, including feeling inexperienced, self-doubt about knowledge of issues at hand, and perceived ramifications for speaking up. One female medical scientist (female, Ph.D., early career) spoke for a number when she said, "Many times I don't know how to make a contribution because I'm quite certain I don't know enough about the issue at hand." Another woman (female, physician, plauteaued) said, it was "too stressful and risky for her to participate" in department decisions, so she talked with the chair individually. This same woman believed her job might be threatened if she spoke up:

> The hardest thing . . . was to be in a department where you couldn't express yourself without getting—feeling that you were jeopardizing your career, and so my personal values . . . I was afraid earlier on that I would. . . . lose my . . . I would get kicked out of the department. . . . [female, physician, plauteaued]

One woman summarized this well, explaining how women have been socialized to think they need to be at a certain level of experience or meet certain qualifications, whereas men do not question their own level of knowledge or experience. "More likely [women] feel they need to be qualified to do something, where, men, in many ways, don't feel that need. . . . they assume they are qualified" (female, physician, plauteaued).

According to some female faculty, in order to be in a position of power and leadership in this authoritarian-style structure, one must dehumanize (female, physician, left) and "out-macho the guys" (female, Ph.D., left). In response, at one school, a group of women faculty met to give each other support. ". . . there is even a secret group of women faculty who met over a year or two to give support and to talk about what was going on, and [a] lot of paranoia that if somebody found out; namely, the chair . . . they'd be the next target. . . ." (female, physician, left).

Discussion

It seems clear that the hierarchy of chairs is a common and well-established structure in medical schools, and it has a significant impact on the faculty work experience and their perception of transparency. Although our research is based entirely on interviews, we heard nearly no comments negating our depiction of the hierarchy of chairs and what were perceived as "chairs for life." There are surely some excellent department chairs who run departments with inclusive and transparent decision making, but this seems largely based on the chair's personal orientation and leadership style. It is more difficult to control for individual chair style variations than it is to focus on a system that allows for little self-regulation. There is greater opportunity for biases and, thus, discrimination to play a role in decision making when there is little transparency.[19,27]

It is obvious that medical schools as bureaucratic organizations need some kind of hierarchy to operate. In such large organizations as academic medical centers, it is not surprising that many faculty feel remote from the upper levels of administration, but it is of much more concern that so many faculty see difficulties with the department chairpersons' managerial styles. The perceived problem with upper administration is that it is "out of touch" with what goes on in the academic trenches, making decisions without adequate transparency, and supporting the power of chairs.

We found that both men and women recognized the hierarchy, but it seemed to have a greater impact on women, creating what may be a real barrier to women's advancement in academic medicine within the hierarchy. Bickel et al. stated that "most women are accustomed to thinking of relationships in terms of support affiliation, whereas men are accustomed to competition and hierarchy."[33] To the extent this is true, this may provide insight into why men discuss the hierarchy in a matter-of-fact tone and experience this as less of an obstacle to advancement. Although we have no direct evidence to connect the hierarchy to women's advancement (that would take a different kind of study), there is little question that women faculty see the hierarchy of chairs with

its open-ended term policy and the reliance on individual chairs' personal orientation for inclusion as both affecting their work lives and their chances for advancement. It may also be that women do not see others like themselves at the top of the hierarchy, which may make it more threatening. Many women see this as "where the lords reign." Although it is not clear how much of the problem with hierarchy is the structure and how much is the incumbent, the indeterminate length of chair appointments creates a calcified structure that is difficult to change (or avoid). As W.I. Thomas' famous sociological dictum states, "Anything that is perceived as real is real in all of its consequences."[34] Here, the insight translates to if women perceive the hierarchy of chairs as a barrier to advancement, it becomes a barrier.

The hierarchy of chairs and the attendant perceived indeterminate term chair policy are not inevitable aspects of academic structure. Some medical schools have performance reviews for chairs, but these do not necessarily affect the length of time a chair serves. University Arts and Science departments typically have a rotating chair, where the chair is appointed or elected to a fixed term (often 3–5 years). In such a system, a chair can be reappointed or reelected, and every few years there is a review to see if this person should or desires to continue as chair. Adopting such a system would go a long way toward reduce the impact of the hierarchy of chairs and make for a more collegial structure. It would also create more openings where women, URM, and younger faculty could advance in academic medicine. Such a policy change could contribute to advancement for women (and by extension, URM and younger faculty) in several ways: (1) women currently perceive hierarchy as a barrier, (2) there are aspects of hierarchy that actually affect women (and men) in ways that are detrimental to their advancement, and (3) the indeterminate length of a chair's tenure allows for less turnover of the chair and, thus, fewer openings.

Most women faculty and many men faculty clearly would prefer a less hierarchical and more collegially oriented organizational structure. The structured hierarchy affects both inclusion and the perception of transparency in decision making. This aligns with our findings that women seek more collaborative work relationships in academic medicine.[30] A flatter, less hierarchical, and more collaborative structure is preferred by most faculty members.

The structured hierarchy, limited inclusion, and lack of decision-making transparency are not the only factors affecting women's advancement in academic medicine. A study by Carr et al.[16] further supports this by showing that female faculty who have experienced gender discrimination in a hierarchical structure report feeling a sense of helplessness to affect change, suggesting that the structure of hierarchy can affect psychosocial feelings and behaviors. Others have found that the hierarchical structure also impeded effective negotiations.[35] A recent survey study in one medical school shows that the top reasons women leave faculty positions include chair/departmental leadership issues, professional advancement, low salary, and personal reasons.[13] The most common reasons men leave include career and professional advancement, low salary, and lack of faculty development/mentoring. Although men and women share some similarities in terms of professional advancement and salary, women in particular express difficulties with chair/departmental issues, which is also reflected in the findings from our study. The current hierarchical structure developed when medicine was populated overwhelmingly by men. The gender composition of medicine and medical schools has changed enormously in the past three decades, and it may be time to reconsider whether some of this structure is optimally functional for the current needs of academic medicine.

This study has limitations. Although we attempted to select representative medical schools, the sample is from only five schools. Moreover, although we endeavored to interview a stratified but widely representative group of faculty, we only interviewed 96 people out of the thousands of academic faculty in the five schools. As noted, we had difficulty locating plateaued men to interview. Finally, hierarchy is only one factor limiting the advancement of women and URMs in academic medicine. How much this factor impacts advancement compared with other factors is still unknown.

Conclusions

Based on interviews with 96 faculty members at five disparate medical schools, we have identified the hierarchy of chairs as a potential barrier for the advancement of women in academic medicine. The fact that chairs are appointed for what appears to be indeterminate tenure creates a number of obstacles for advancement, especially given the calcified academic structure, including problems with inclusion and transparency in decision making and, given the infrequent turnover in chairs, a bottleneck for advancement. Women faculty seem more affected by this hierarchical structure than men, and addressing this may help the advancement of women in academic medicine.

Acknowledgments

We gratefully acknowledge the critical funding support of the Josiah Macy, Jr. Foundation and the supplemental funds to support data analysis provided by Office of Public Health and Science, Office on Women's Health, and Office on Minority Health; National Institutes of Health, Office on Research on Women's Health; Agency for Healthcare Research and Quality; Centers of Disease Control and Prevention; and Health Resources and Services Administration.

Thanks as well to Marianne McPherson and Meg Lovejoy for assistance with data coding and to the reviewers for their helpful comments. We are indebted to the medical faculty who generously shared their experiences in the interviews.

Disclosure Statement

The authors have no conflicts of interest to report.

References

1. Association of American Medical Colleges. Women in U.S. academic medicine: Statistics and benchmarking report 2007–2008. Available at www.aamc.org/members/wim/statistics/stats08/stats_report.pdf Accessed August 13, 2009.
2. Association of American Medical Colleges. An overview of women in U.S. academic medicine, 2006–2007. Available at www.aamc.org/members/wim/statistics/stats07/start.htm Accessed April 15, 2009.
3. Association of American Medical Colleges. Women in U.S. academic medicine: Statistics and benchmarking report 2007–2008. Table 4. Distribution of women faculty by

department, rank, and degree. Available at www.aamc.org/members/wim/statistics/stats08/table04.pdf Accessed August 13, 2009.
4. Association of American Medical Colleges. Women in U.S. academic medicine: Statistics and benchmarking report 2007–2008. Table 3 Distribution of faculty by department, rank, and gender. Available at www.aamc.org/members/wim/statistics/stats08/table03.pdf Accessed August 13, 2009.
5. Association of American Medical Colleges. Women in U.S. academic medicine: Statistics and benchmarking report 2007–2008. U.S. medical school faculty distribution by gender and rank. Available at www.aamc.org/members/wim/statistics/stats08/start.htm Accessed August 13, 2009.
6. Association of American Medical Colleges. Women in U.S. academic medicine: Statistics and benchmarking report 2007–2008. Table 9. 2007 Benchmarking—Division/section chiefs and department chairs. Available at www.aamc.org/members/wim/statistics/stats08/table09.pdf Accessed August 13, 2009.
7. Association of American Medical Colleges. Women in U.S. academic medicine: Statistics and benchmarking report 2007–2008. Table 11 Women deans and interim deans, October 2008. Available at www.aamc.org/members/wim/statistics/stats08/table11.pdf Accessed August 13, 2009.
8. Ness RB, Ukoli F, Hunt S, et al. Salary equity among male and female internists in Pennsylvania. Ann Intern Med 2000;133:104–110.
9. Tesch BJ, Wood HM, Helwig AL, Nattinger AB. Promotion of women physicians in academic medicine. Glass ceiling or sticky floor? JAMA 1995;273:1022–1025.
10. Weeks WB, Wallace AE. Race and gender differences in general internists' annual incomes. J Gen Intern Med 2006;21:1167–1171.
11. Ash AS, Carr PL, Goldstein R, Friedman RH. Compensation and advancement of women in academic medicine: Is there equity? Ann Intern Med 2004;141:205–212.
12. Buckley LM, Sanders K, Shih M, Kallar S, Hampton C. Obstacles to promotion? Values of women faculty about career success and recognition. Committee on the Status of Women and Minorities, Virginia Commonwealth University, Medical College of Virginia Campus. Acad Med 2000;75:283–288.
13. Cropsey KL, Masho SW, Shiang R, Sikka V, Kornstein SG, Hampton CL. Why do faculty leave? Reasons for attrition of women and minority faculty from a medical school: Four-year results. J Womens Health 2008;17:1111–1118.
14. Wright AL, Schwindt LA, Bassford TL, et al. Gender differences in academic advancement: Patterns, causes, and potential solutions in one U.S. college of medicine. Acad Med 2003;78:500–508.
15. Carr PL, Ash AS, Friedman RH, et al. Faculty perceptions of gender discrimination and sexual harassment in academic medicine. Ann Intern Med 2000;132:889–896.
16. Carr PL, Szalacha L, Barnett R, Caswell C, Inui T. A "ton of feathers": Gender discrimination in academic medical careers and how to manage it. J Womens Health 2003;12:1009–1018.
17. Oakley JG. Gender-based barriers to senior management positions: Understanding the scarcity of female CEOs. Business Ethics 2000;27:321–334.
18. Gamba M, Kleiner BH. The old boys' network today. Int J Sociol Social Policy 2001;21:101–107.
19. Reskin BF. Getting it right: Sex and race inequality in work organizations. Annu Rev Sociol 2000;26:707–709.
20. Kanter RM. The impact of hierarchical structures on the work behavior of women and men. In: Myers KA, Anderson CD, Risman BJ, eds. Feminist foundations: Towards transforming sociology. Thousand Oaks, CA: Sage, 1977:259–277.
21. Kanter RM. Men and women of the corporation. New York: Basic Books, 1977.
22. Kanter RM. Some effects of proportions on group life: Skewed sex ratios and responses to token women. Am J Sociol 1977;82:965–990.
23. Pfeffer J, Davis-Blake A. The effect on the proportion of women on salaries: The case of college administrators. Admin Sci Q 1987;32:1–24.
24. Reskin BF. Sex segregation in the workplace. Annu Rev Sociol 1993;19:241–270.
25. Cohen L, Broschak J, Haveman H. And then there were more? The effect of organizational sex composition on the hiring and promotion of managers. Am Sociol Rev 1998;63:711–727.
26. Bielby WT. Minimizing workplace gender and racial bias. Contemp Sociol 2000;29:120–129.
27. Reskin BF. The proximate causes of employment discrimination. Contemp Sociol 2000;29:319–328.
28. Carnes M, Morrissey C, Geller SE. Women's health and women's leadership in academic medicine: Hitting the same glass ceiling? J Womens Health 2008;17:1453–1462.
29. Carr P, Pololi L, Knight S, Conrad P. Collaborating in academic medicine: Reflections on gender and advancement. Acad Med 2009;84:1447–1453.
30. Pololi L, Conrad P, Knight S, Carr P. A study of the relational aspects of the culture of academic medicine. Acad Med 2009;84:106–114.
31. Charmaz K. Constructing grounded theory: A practical guide through qualitative analysis. Thousand Oaks, CA: Sage, 2006.
32. Glaser B, Strauss AL. The discovery of grounded theory: Strategies for qualitative research. Chicago: Aldine, 1967.
33. Bickel J, Wara D, Atkinson BF, et al. Increasing women's leadership in academic medicine: Report of the AAMC Project Implementation Committee. Acad Med 2002;77:1043–1061.
34. Thomas WI. Social behavior and personality. New York: New York Research Council, 1955.
35. Sarfaty S, Kolb D, Barnett R, et al. Negotiation in academic medicine: A necessary career skill. J Womens Health 2007;16:235–244.

Address correspondence to:
Peter Conrad, Ph.D.
Brandeis University
415 South Street, Mailstop 071
Waltham, MA 02454

E-mail: conrad@brandeis.edu

CLINICAL TRIALS *RECRUITMENT AND RETENTION* *Clinical Trials* 2010; **7**: 90–99

Impact of supplemental site grants to increase African American accrual for the Selenium and Vitamin E Cancer Prevention Trial

Elise D Cook[a], Kathryn B Arnold[b], John A Hermos[c], Worta McCaskill-Stevens[d], Sarah Moody-Thomas[e], Jeffrey L Probstfield[f], Sandra J Hamilton[g], Russell D Campbell[h], Karen B Anderson[i] and Lori M Minasian[d]

Background African American accrual to prevention trials at rates representative of the disease burden experienced by this population requires additional resources and focused efforts.

Purpose To describe the rationale, context, and criteria for selection of sites that received Minority Recruitment Enhancement Grants (MREGs) to increase African American recruitment to the Selenium and Vitamin E Cancer Prevention Trial (SELECT). To determine if African American accrual was higher among the 15 MREG sites when compared with similar nonawarded sites.

Methods Changes in African American accrual at sites that received MREGs are compared with changes in a group of 15, frequency-matched, nonawarded sites using a quasi-experimental, *post hoc* analysis. Successful and unsuccessful recruitment strategies reported by the MREG sites are described.

Results The increased number of African American participants accrued per month at MREG sites post-funding was higher than the change at comparison sites by a factor of 3.38 ($p = 0.004$, 95% CI: 1.51–7.57). An estimated 602 additional African American participants were recruited at MREG sites due to MREG funding, contributing to the overall 14.9% African American recruitment. Successful recruitment strategies most reported by MREG sites included increasing staff, transportation resources, recruiting through the media, mailings, and prostate cancer screening clinics during off-hours.

Limitations Comparison sites were chosen retrospectively, not by randomization. Although comparison sites were selected to be similar to MREG sites with regard to potential confounding factors, it is possible that unknown factors could have biased results. Cost-effective analyses were not conducted.

Conclusions MREG sites increased African American accrual in the post-funding period more than comparison sites, indicating MREG funding enhanced the sites' abilities to accrue African American participants. Targeted grants early in the accrual period may be a useful multi-site intervention to increase African American accrual for a prevention study where adequate African American representation is essential. *Clinical Trials* 2010; **7**: 90–99. http://ctj.sagepub.com

[a]MD Anderson Cancer Center, University of Texas, Houston, TX, USA, [b]Southwest Oncology Group Statistical Center, Seattle, WA, USA, [c]VA Cooperative Studies Program, VA Boston Healthcare System, Boston, MA, USA, [d]National Cancer Institute, Division of Cancer Prevention, NIH, Bethesda, MD, USA, [e]LSU, Health Sciences Center, School of Public Health, New Orleans, LA, USA, [f]University of Washington, Seattle, WA, USA, [g]Morehouse School of Medicine, Atlanta, GA, USA, [h]Fred Hutchinson Cancer Research Center, Seattle, WA, USA, [i]Cancer Research And Biostatistics, Seattle, WA, USA
Author for correspondence: Elise D Cook, MD Anderson Cancer Center, University of Texas, Clinical Cancer Prevention, Unit 1360, P.O. Box 301439, Houston, TX 77230-1439, USA. E-mail: edcook@mdanderson.org

10.1177/1740774509357227

Introduction and background

Prostate cancer is a major health problem, second only to nonmelanoma skin cancer as the most common cancer in men in the United States. Furthermore, the rate of prostate cancer is higher and the mean age-of-onset is younger in African Americans than in non-African Americans [1–3]. Prostate cancer prevention trials (PCPTs) provide an important opportunity to test interventions that might reduce the burden of the disease [4]. However, unambiguous application of the results of such trials to African Americans depends upon adequate participation by African American men.

The PCPT was the first large, cooperative group randomized trial for the prevention of prostate cancer in healthy men [5]. The randomized participant goal for African American men was set at 8% to mirror the estimate of African American men aged $\geq$55 years in the US population. Only 4% African American men were randomized to PCPT during the 3-year enrollment period. Efforts to enhance minority participation in PCPT were not initiated until 1 year after the study was activated. In addition, about two-thirds of the overall accrual goal was met in the first year of recruitment, so any enhanced African American enrollment after that time could have had only a modest overall effect.

The PCPT minority recruitment experience suggested that successful recruitment of African American men into a PCPT requires recruitment efforts to be initiated at trial activation, infrastructure provided to support minority recruitment and a long-term commitment from funding agencies [6–8]. Known barriers commonly cited to impede minority recruitment must also be addressed, such as the attitudes, knowledge, and beliefs of potential minority recruits and their referring clinicians, as well as trial designs and costs [9–17].

Minority recruitment lessons learned from PCPT were applied to Selenium and Vitamin E Cancer Prevention Trial (SELECT), the next large, cooperative group prostate cancer prevention study. SELECT was designed to evaluate the effect of selenium and vitamin E on the incidence of prostate cancer without changing the clinician's practice of prostate cancer screening. The SELECT overall randomization goal was 32,400 healthy men, with 6480 men enrolled every year for 5 years. The study had a pre-established goal of 24% overall minority representation: 20% African American, 3% Hispanic, and 1% Asian/Pacific Islander [18].

Prior to trial initiation, SELECT took several steps to enhance recruitment of African American and other racial/ethnic minorities. First, the eligibility criteria were expanded. Second, sites with prior success in minority recruitment were sought, and the Department of Veteran Affairs Cooperative Studies Program was included as a SELECT affiliate because of its strong track record in minority recruitment [19]. Third, SELECT developed a national infrastructure to support minority recruitment. Fourth, SELECT provided additional funds in the form of Minority Recruitment Enhancement Grants (MREGs) to sites with the potential to increase minority enrollment [20].

Five months after study activation, SELECT was enrolling participants at nearly twice the planned rate. Although Hispanic and Asian/Pacific Islander recruitment met or exceeded the targeted enrollment rates, while African American recruitment was much lower than anticipated, it appeared that the overall enrollment goal would be achieved in less than the planned 5 years. SELECT investigators had to respond quickly to boost African American participation. Due to the projected shortened recruitment period, a targeted yet flexible intervention was needed to increase African American participation.

Two MREG requests for applications were distributed during the accrual period. The MREG was a $50,000 site grant, provided by the National Cancer Institute and designed to increase minority enrollment by enhancing recruitment strategies at sites with the potential to increase minority recruitment. An additional purpose of the MREG was to document the success of recruitment strategies used by sites for the benefit of future studies. While all SELECT sites could apply for an MREG, applicant sites were required to demonstrate the ability to recruit minorities and/or provide evidence of access to large numbers of minority men. The MREG requests for applications focused on all minority recruitment, but applications were scored higher if African Americans were the targeted population. Sites were notified of awards within 1–3 months following receipt of their application.

The National Cancer Institute provided funding to SELECT for 11 MREGs that were awarded in mid-2002. Only 32 sites out of all 427 SELECT sites applied, fewer than, had been anticipated. At the time of this initial funding, total recruitment was 10,500 and African American participation was at 9.8%. One year later, when overall SELECT enrollment reached 20,000 and African American participation was 12.4%, the National Cancer Institute allocated additional funding to issue another request for applications. As a result of this second round of competition in mid-2003, seven existing MREG sites and four new sites were awarded MREGs. In summary, 15 SELECT sites received a combined $1.1 million in grants over 2 years to increase African American and other minority participation (Table 1).

Table 1 Summary of MREG awards to SELECT sites

	Funding dates		
Site #	1st MREG	2nd MREG	Total funds
1[a]	July 2002	July 2003	$100,000
2[a]	July 2002	July 2003	$100,000
3	July 2002	July 2003	$100,000
4	July 2002	July 2003	$100,000
5	July 2002	July 2003	$100,000
6	July 2002		$50,000
7[b]	July 2002		$50,000
8	July 2002		$50,000
9	August 2002	August 2003	$100,000
10	September 2002	September 2003	$100,000
11[a]	October 2002		$50,000
12		April 2003	$50,000
13		April 2003	$50,000
14		August 2003	$50,000
15[a]		August 2003	$50,000

[a]Site had a large percentage of African American recruits pre-funding. MREG was awarded to maintain already high African American recruitment. [b]Very large site with access to large numbers of Hispanic and African American recruits. MREG was awarded to enhance both African American and Hispanic recruits.

Recruitment for SELECT ended in June 2004, 9 months after the last site received notification of a 2003 MREG award. SELECT accrued 35,533 participants, of whom 14.9% were African American. Although African American enrollment was less than the 20% goal, it was three times that seen in PCPT and the largest percentage of African Americans ever accrued to a cancer prevention trial.

This article describes the rationale, context, and criteria for the selection of sites that received additional funding to enhance African American accrual in the SELECT, a prostate cancer prevention study. The changes in African American recruitment at these sites compared with similar, nonawarded sites were evaluated in a quasi-experimental, *post hoc* analysis. Descriptions of how sites used these funds and what benefits and limitations these sites reported are also presented.

Methods

Quantitative study design

To determine whether MREG funding impacted African American enrollment in SELECT, the change in African American accrual rates before and after sites received funding was evaluated by comparing MREG sites with similar non-MREG sites in a *post hoc* analysis. In this quasi-experimental study design, the intervention (MREG) was not randomly assigned. The number and percentage of African American participants accrued before and after MREG sites received funding was measured. Comparisons are presented as the change in accrual rates and odds ratios for African American versus non-African American recruitment. For the purposes of this analysis, a participant's race was determined by self-report at randomization and was recorded for SELECT in compliance with current National Cancer Institute standards. If a participant self-reported as African American, regardless of other race affiliations or ethnicity, he was considered African American.

MREG sites were frequency matched to a set of non-MREG sites with similar early enrollment patterns to obtain a comparison group of sites with similar potential for African American accrual. The matching process involved consideration of two factors. The rate at which the sites were already randomizing African American participants to SELECT early in the enrollment period (prior to receipt of any MREG funds) was a strong potential confounder. Therefore, the primary matching factor was based on the percentage of African American participants accrued as of May 31, 2002, prior to receipt of MREG funding at any site. The total number of recruits prior to May 31, 2002, was the second factor in choosing comparison sites. An exception is Site 15, which started randomizing participants after May 31, 2002; total and percentage African American accrual as of June 30, 2003, was used instead. Each comparison site was assigned a hypothetical funding date based on the date the similar, matched, MREG site received an MREG award. This site-specific date separated the pre- and post-funding periods for purposes of

Table 2 Characteristics of SELECT sites awarded MREGs and non-MREG comparison sites

Characteristic	MREG sites $N=15$	Comparison sites $N=15$
Percentage African American participants accrued as of May 31, 2002 (%)		
0–9	3	3
10–24	6	6
25–49	3	3
50–74	0	3
75–100	3	0
Number of participants accrued as of May 31, 2002		
0–24	2	4
25–49	3	2
50–99	4	6
100–199	5	3
200+	1	0
Type of site[a]		
Academic Center	6	5
CCOP	4	1
Cancer Program	2	3
VACSP	2	4
Community Health Center	1	0
Private practice	0	2
Existing foundation for African American accrual pre-funding[b]		
Yes	14	8
No	1	5
Unknown	0	2
Ever utilized non-MREG SELECT resources to recruit African American participants[c]		
Yes	5	2
No	10	11
Unknown	0	2

[a]Type of site. *Academic Center*: Facility is also involved in higher education and research. *CCOP, Community Clinical Oncology Program*: A large network that allows community physicians to participate in NCI-sponsored clinical cancer trials.*Cancer Program*: Facility that has a cancer program, may or may not be an academic center, not an NCI-designated Cancer Center or Comprehensive Cancer Center. *VACSP, Department of Veteran Affairs Cooperative Studies Program*: Department of Veterans Affairs provides health care to eligible veterans; VACSP coordinates large multisite research projects. *Community Health Center:* Facility that provides public health services. *MBCCOP, Minority-Based Community Clinical Oncology Program*: allows racial and ethnic minority cancer patients to have access to quality medical care in their own communities. [b]For example, sites had established relationships in the minority community. [c]For example, sites participated in national SELECT minority recruitment initiatives.

the analyses. A summary of MREG and comparison site characteristics is shown in Table 2.

As a whole, the MREG and comparison sites had similar characteristics. Each group included Academic Centers, VA Cooperative Study sites, Community Clinical Oncology Program sites, and Cancer Programs. There were no Community Health Centers among the comparison sites and no private practices among the MREG sites. Most sites in both groups had pre-existing foundations for accruing African Americans and had never used non-MREG SELECT resources to recruit African American participants to SELECT (Table 2).

Quantitative analysis

The outcomes of interest are (1) the accrual rate of African Americans, that is, the number of African Americans enrolled per month and (2) the probability that a randomized participant is African American. Although the event of interest (whether the participant randomized is African American) occurs at the participant level within each site, MREG is an intervention applied at the site level, giving rise to clustered data. Mixed effects regression models were used with the individual as the primary unit of analysis and site-specific random

Table 3 Monthly accrual of African Americans and all participants for SELECT sites awarded MREGs and non-MREG comparison sites

	Monthly African American accrual				Monthly accrual for all participants			
	MREG sites[a]		Comparison sites[a]		MREG sites		Comparison sites	
Site #	Pre-funding[b] *N* (%)	Post-funding[c] *N* (%)	Pre-funding *N* (%)	Post-funding *N* (%)	Pre-funding *N*	Post-funding *N*	Pre-funding *N*	Post-funding *N*
1	5.0 (40)	5.8 (52)	3.7 (45)	1.5 (22)	12.5	11.0	8.2	6.6
2	2.1 (75)	6.8 (91)	1.3 (50)	0.6 (31)	2.8	7.4	2.7	1.8
3	1.5 (18)	0.5 (27)	0.9 (19)	2.3 (52)	7.7	1.8	4.6	4.4
4	3.3 (21)	2.8 (29)	1.5 (21)	1.0 (39)	15.4	9.7	7.1	2.6
5	0.4 (20)	2.4 (40)	0.3 (15)	0.4 (41)	2.0	6.0	1.9	1.0
6	0.1 (12)	0.6 (28)	0.1 (14)	0.0 (0)	0.8	2.2	0.7	0.1
7	1.4 (4)	2.4 (4)	1.3 (6)	0.0 (12)	29.0	52.5	22.0	0.3
8	6.1 (37)	2.7 (65)	10.1 (42)	6.5 (39)	16.2	4.1	23.7	16.5
9	1.6 (15)	2.3 (29)	1.1 (13)	0.2 (11)	10.7	7.9	8.2	2.0
10	1.8 (33)	1.4 (43)	1.5 (32)	0.1 (60)	5.3	3.3	4.8	0.2
11	2.5 (94)	1.5 (91)	1.1 (37)	0.4 (32)	2.6	1.6	2.9	1.3
12	1.4 (22)	0.3 (11)	0.7 (17)	0.0 (0)	6.3	2.9	3.7	0.3
13	0.2 (7)	1.3 (28)	0.1 (3)	0.0 (0)	2.4	4.7	1.7	0.5
14	1.0 (10)	3.6 (28)	0.2 (4)	0.3 (18)	10.5	12.8	4.2	1.5
15	7.5 (96)	8.5 (100)	3.4 (62)	3.5 (66)	7.8	8.5	5.5	5.2
Mean[d]	2.4 (34)	2.9 (44)	1.8 (25)	1.1 (28)	8.8	9.1	6.8	3.0

[a]Similar MREG and comparison sites are shown on the same line. [b]Pre-funding is the site-specific period starting from the month of the first randomization at the site through the month prior to the initiation of MREG funding at that site, as shown in Table 1. The pre-funding period for each comparison site ends at the same time as the similar MREG site; comparison sites may have initiated accrual in different months than the MREG sites. [c]Post-funding is the site-specific period starting from the first month of MREG funding at that site through the end of SELECT accrual on June 24, 2004. The post-funding period for each comparison site is the same as the similar MREG site. [d]To estimate the average effect size across sites, means were calculated as follows: mean ''*N*'' is the sum of the monthly number of participants accrued by each site divided by 15; mean ''%'' is the sum of the percentage African American participants accrued by each site divided by 15.

effects were used to account for the correlation between participants accrued to the same site.

For the tabular data presentation, the number of participants accrued per month was the total enrollment at a site for the time period, pre- or post-funding, divided by the number of months the site had been accruing participants within the specified time period (Table 3). Percentage African American accrual was determined by the number of African Americans accrued divided by the total number of participants accrued, for each time period per site. Mean number of participants was the sum of the monthly number of participants accrued by each site divided by 15. Mean percentage African American accrual was calculated in the same fashion, to give an approximate average effect size across sites by giving each site equal weight.

All analyses were performed using PROC NLMIXED in SAS 9.1 (SAS Institute Inc, Cary, North Carolina). To predict the number of African American participants accrued per month, the number of African Americans recruited during the entire pre- or post-funding period was used as the response and a mixed effects model assuming a Poisson distribution was fitted. The response was modeled employing a log link function and included an offset for number of months of enrollment. To estimate the odds of randomizing African American participants, a mixed effects model was used to predict the probability that a randomized participant was African American in the pre- or post-funding time periods, using a Bernoulli distribution with a logit link function. Both models included fixed effects for MREG status, time (pre- or post-funding), and an interaction term for time and MREG status. Although these were not matched analyses, the frequency matched factors were included in the models: the percentage of African American participants and number of participants accrued through May 2002. The models included random site effects for pre- and post-funding, assumed to be normally distributed with mean 0 with an unstructured covariance matrix. A test of the interaction term was conducted to determine if the change in African American accrual patterns from pre- to post-funding were different when comparing the MREG sites with the comparison sites.

The expected number of participants accrued to MREG sites in the absence of MREG funding during the post-funding period was estimated using parameters from the Poisson model by setting the

MREG term to 0. The difference between this estimate and the actual African American enrollment post-funding at MREG sites provides an approximation of the additional number of African American participants accrued due to MREG funding.

Qualitative analyses

All MREG sites were required to submit monthly status reports during the period covered by MREG funding. These reports included whether or not specified minority recruitment strategies were successful. A successful strategy was defined for the sites as one that resulted in at least one randomization at a site in a given month, and an unsuccessful strategy was defined as one that had been used by the site but resulted in no randomizations that month. Successful and unsuccessful strategies were ranked by the number of times sites reported them.

In an effort to assess the overall impact of the MREG on randomizations and how the process could be enhanced, the 11 sites receiving the second round of MREG funding were asked to complete a final MREG summary report. The survey was comprised of four open-ended questions and one Likert item. These items were designed to probe the sites' experience with the MREG process by identifying institutional barriers to efficient use of MREG funds; evaluating the effectiveness of time spent on MREG-related activities; capturing lessons learned from the MREG experience; and providing an opportunity to present any relevant but previously unsolicited information.

Results

Quantitative findings

In general, overall enrollment for the 15 MREG and 15 comparison sites was steady throughout most of the recruitment period. The MREG sites enrolled 4507 total participants including 1266 (28%) African Americans; the comparison sites accrued 4025 participants of which 911 (23%) were African Americans. The rate of African American accrual for the MREG sites increased more than the comparison sites near the end of recruitment (Figure 1).

Monthly accrual data for the MREG and comparison sites are presented in Table 3. To allow for an evaluation of the enrollment differences between MREG and comparison sites, monthly randomization data are presented for African American and all participants by site, using means to provide an estimate of the average effect size across sites. MREG and similar comparison sites are presented side-by-side to best demonstrate the similarity of their baseline African American recruitment percentages. Other than the matched comparison site for MREG Site 11, these percentages can be seen to compare quite closely. Mean monthly African American accrual for MREG sites increased from 2.4 to 2.9 African American participants per month, while the comparison sites experienced a decrease from 1.8 to 1.1 African American participants per month. Mean percentage African American accrual across MREG sites increased from 34% to 44%; the comparison sites increased only modestly, from 25% to 28%. The mean number of participants accrued by MREG sites increased slightly, from 8.8 to 9.1 participants per month, and the comparison sites showed a decrease, from 6.8 to 3.0 participants per month. Thus, despite a minimal increase in the percentage of African American recruits between pre- and post-funding periods among the comparison sites, the low overall recruitment at comparison sites resulted in a diminished rate of African American recruitment per month.

Based on the results from the Poisson regression model, there was no evidence of a change in the monthly number of African Americans accrued to MREG sites post-funding (rate ratio = 1.28 $p = 0.84$ 95% CI: 0.47–3.51). The number of African Americans randomized to the comparison sites per month decreased overall by a factor of 0.38 ($p = 0.003$ 95% CI: 0.21–0.69) in the post-funding time period. The change in monthly African American accrual rate for MREG sites is 3.38 times the monthly change in accrual rate for comparison sites ($p = 0.004$ 95% CI: 1.51–7.57). This ratio indicates that while the MREG sites maintained the same levels of African American enrollment, the comparison sites' rates of enrollment of African Americans declined.

Results from the Bernoulli regression model showed the proportion of African American participants at the MREG sites increased after the receipt of funding. The odds for African American enrollment post-funding was 1.99 ($p = 0.08$ 95% CI: 0.91–4.32) times the odds pre-funding. In contrast at the comparison sites, the proportion of African American accrued did not increase (OR = 1.07 $p = 0.77$ 95% CI: 0.68–1.69). Comparing changes for MREG to non-MREG sites, the relative odds ratio was 1.86 ($p = 0.048$ 95% CI: 1.01–3.43), demonstrating that the MREG sites increased their odds of accruing African American participants post- versus pre-funding while the comparison sites did not.

MREG sites accrued 850 African American participants after receiving MREG funding. Based on the Poisson model parameters, the expected number of

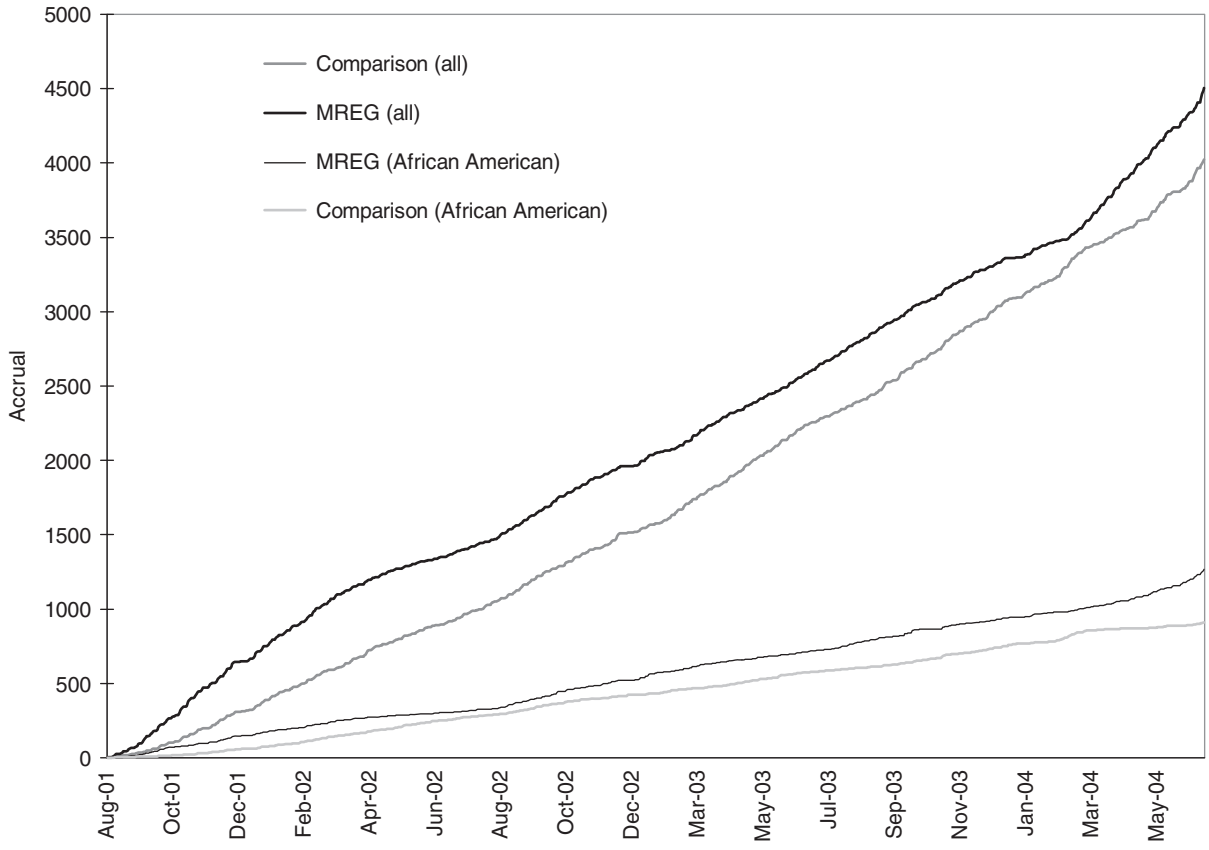


Figure 1 Cumulative accrual for sites awarded MREGs and non-MREG comparison sites for all participants and African Americans only. MREG sites ($n=15$) accrued 4507 participants with 1266 African Americans. Non-MREG comparison sites ($n=15$) accrued 4025 participants with 911 African Americans

African American participants these sites would have accrued in the absence of MREG funding is 248. The difference between these values, 602, is the approximate additional number of African American participants accrued at MREG sites due to MREG funding.

Qualitative findings

All MREG sites submitted monthly reports including minority recruitment strategies attempted during the funding period. Recruitment strategies that most frequently resulted in randomizations included: (1) additional staff time; (2) providing resources such as transportation and parking support, minority recruitment materials, and refreshments for recruitment meetings; (3) recruiting through the media; (4) mass mailings; and (5) prostate cancer screening clinics during off-hours. Strategies that were reported to result in very few randomizations and which some sites rated as unsuccessful included: (1) publicizing SELECT at health fairs, churches, barbershops, laundromats, and grocery stores; (2) local spokespersons; and (3) networking with clinics and community leaders. Some participants were recruited by affiliation with PCPT. The following three unsuccessful strategies included: (1) publicizing SELECT as a link on a local website, (2) targeting women's groups, and (3) publicizing the National Cancer Institute's website for cancer prevention information.

Eight of 11 eligible sites completed final MREG summary reports at the end of the second MREG funding cycle. Of these sites, 88% rated the time they spent on MREG-related activities as 'effective' or 'very effective'. Eighty-eight percentage of these sites also recommended that future studies provide additional funding to enhance African American recruitment before trial activation. The most common benefits listed by these sites included extra staff time (including evenings and weekends) to focus on African American recruitment and assist with planning, outreach, and hosting activities and funds for advertising and mass mailings. All sites reported hiring additional staff, usually a minority

outreach coordinator, and most sites increased existing staff time for minority recruitment. Other activities supported by the MREG included transportation and parking support, minority recruitment materials, recruitment advertisements in local media, food and supplies for recruitment meetings and postage for mass mailings.

MREG sites reported barriers that included funding and staffing delays, the absence of staff during summer vacation season when the first MREG became available, minority recruiter illness, and possible participant distrust of clinical trials. Some sites reported that it took additional staff time to recruit African American men. Sites also reported screening men who were not enrolled because they lived outside any SELECT study site area or they were African American men old enough for prostate screening but too young to qualify for SELECT.

Similar information on strategies used to enhance African American recruitment was not requested from the 15 comparison sites. Hence, it was not possible to determine whether the strategies reported by MREG sites, or their purported successes or failures, were similar or different from those possibly employed by comparison sites.

Discussion

We have presented qualitative and quantitative data that explore the impact of MREGs on African American recruitment in the SELECT, a large, multisite prostate cancer prevention study. Although SELECT did not meet the target of 20% for African American enrollment, the MREG sites contributed to an increase in African American randomizations. When compared to 15 sites that did not apply for or receive MREG funding, matched in a *post hoc*, quasi-experimental design, MREG sites had statistically significant increases in African Americans accrued per month and higher odds of accruing African American participants when comparing post- versus pre-funding periods. Both differences are statistically significant, indicating the MREG sites were successful in increasing African American randomizations after receiving the grants.

In this quasi-experimental analysis, the choice of comparison sites imparts an important limitation to these findings. The comparison sites had not applied for MREG funding, and there are no measures of their actual potential and motivation for either maintaining or expanding African American recruitment in the latter stages of SELECT recruitment. If a different mix of comparison sites had been chosen on a primary factor other than on pre-funding frequency of African American recruitment, differences in African American recruitment rates between MREG and non-MREG sites might have been more or less pronounced. Another limitation is that the amount of funding each site received was not considered, only the event of ever having received MREG funding.

A cost effectiveness analysis of this intervention was not conducted. However, it is estimated that MREG funding resulted in 602 additional African American participants. If the only outcome of importance from the MREG is African American enrollment, then the cost per additional African American participant accrued is approximately $1827, contrasted to the $1000 SELECT paid to the sites for each randomization. The intervention was expensive but effective.

Simply providing an additional $827 to SELECT sites for each African American man randomized may have increased the number of African American men enrolled to SELECT but probably would not have increased the overall recruitment of men to SELECT as seen with the MREG awards. Additionally, some sites would not have the access or resources to recruit from the African American population, regardless of the financial incentive to do so.

There were other potential benefits to the sites who received these grants that are unmeasured but important to the conduct of SELECT and future trials. These potential benefits could include: increasing overall randomizations by MREG sites; gaining a greater presence and support in the targeted communities; providing health education and clinical trials information to the community; and increasing a recruitment base for future studies.

The impact of the MREG on African American and total recruitment varied among the funded sites. Although these sites were chosen for their ability to recruit African American participants, not all MREG sites increased African American randomizations. For example, Site 12 was unable to hire a minority recruiter as planned, which contributed to their decrease in overall and African American enrollment after receipt of MREG funding. Some sites, such as Site 15, with high initial African American randomizations, used the funds to increase their overall enrollment rates while maintaining that of African Americans. Site 4 had a decline in overall enrollment rates just prior to receipt of MREG funding. They reported that their existing financial resources were being depleted, that the current staff was unable to handle additional volume, and that they were unable to continue their recruitment efforts at the same level in the absence of assistance. Had the MREG not been available, Site 4 would probably have made a reduced contribution to African American

participation. Additionally, Site 7, the largest MREG site, improved only Hispanic recruitment after receiving MREG funding, even though its stated intent was to increase both Hispanic and African American participation.

There are several potential sources of bias for this study. First, the intervention sites were not randomly selected, but were chosen via an application process where they had to demonstrate the potential and ability to recruit African American participants to SELECT. Some sites may have had existing minority recruitment programs in place; other sites may have been initiating African American recruitment efforts. Sites may have been experiencing a decrease in randomizations, such that MREG funding allowed them to continue African American enrollment at their preliminary rate, rather than increasing accrual; other sites may have been able to use funds to boost existing African American randomizations.

Varying degrees of success and failure were seen with some strategies, including publicizing SELECT at health fairs, networking with clinics, working with community leaders, and recruiting through churches. These variations speak to site and community differences and the intricacies of recruitment methodologies. Although we recommend the successful strategies used in SELECT, we realize that one size does not fit all, and sites need to have flexibility as to which strategies they pursue.

Although application for the first MREG award was open to all SELECT sites, only 32 SELECT sites ($< 10\%$) responded to the first request for applications. MREG funding was implemented after the trial was open and when sites were actively enrolling participants, overall at a higher than anticipated rate. This early success in overall accrual may have contributed to the fewer than expected number of sites applying for an MREG. Most sites would have established staff assignments to SELECT and budgeted funds and time commitments prior to the first MREG announcement. These site staff may have perceived that their site infrastructure and accrual goals could not accommodate the additional work involved in enhancing minority recruitment. Site staff may not have believed that they were able to increase minority enrollment to SELECT even with additional funding due to lack of access to the African American population or workload issues. A number of sites experienced delays in gaining access to MREG funds due to impediments within their own institutions. This resulted in further hiring and implementation delays. These factors lend further support for initiating minority recruitment strategies at trial inception and implementing them at the onset of randomization.

National African American minority recruitment strategies promoting prostate screening and SELECT may have augmented African American accrual for both MREG and non-MREG sites. For example, all sites were strongly encouraged to enroll African American men to SELECT, not just MREG sites. Three minority accrual workshops were conducted to increase African American enrollment to SELECT. The participating sites included both MREG and non-MREG sites with the potential to enhance African American recruitment. These workshops provided a forum for sites to exchange ideas form mentoring relationships and discuss strategies. Other nationwide minority recruitment strategies included 'SELECT Sunday', a faith-based strategy initiated in November 2003; African American media personalities participated in limited media spots promoting SELECT; and a barbershop initiative that preceded the release of the movie Barbershop 2, which opened February 6, 2004. Annual events surrounding Prostate Cancer Awareness Month, Minority Cancer Awareness Week, and Fathers Day also had the potential to boost enrollment to SELECT, and some sites tailored these events to enhance African American recruitment.

Future researchers considering the use of targeted funding to boost minority participation are recommended to make funds available and plan minority recruitment strategies prior to trial activation. A study should be ready to implement strategies before recruitment opens, choose recipient sites wisely, minimize delays incurred with hiring additional staff, and allow flexibility in the use of funds so sites can tailor interventions to their own needs. The information and strategies presented here should help guide future large-scale prevention and treatment studies, where recruitment of sufficient numbers of African American participants is necessary ethically and as a practical means to generalize results to this population.

Acknowledgments

We are indebted to the following: SWOG Chairman Emeritus, Charles A. Coltman, MD; NCI, Division of Cancer Prevention, Associate Director for Clinical Research, Leslie Ford, MD; Study Coordinators, Eric A. Klein, MD, Scott M. Lippman, MD, and Ian M. Thompson, Jr, MD, J. Michael Gaziano, MD, MPH, Department of Veterans Affairs; Biostatisticians, John J. Crowley, PhD and Phyllis J. Goodman, MS; Project Manager, Jo Ann Hartline, MS; the SELECT Retention and Adherence Committee and the SELECT Minority and Medically Underserved Subcommittee for their valuable assistance during the recruitment period of the trial; Margaret Pepe, PhD, and Dale F. McLerran, MS, for their statistical support through the

Biostatistics Shared Resource at the Fred Hutchinson Cancer Research Center; Jason M. Harris-Talley for data acquisition, Alma Garcia for data entry assistance, and Tamesha Thomas for assistance in preparing the article. We would also like to acknowledge the investigators and staff at 427 SELECT study sites for completing accrual to SELECT well ahead of schedule.

This study is supported in part by grants 5-U10-CA-37429 and P30-CA015704 from the National Cancer Institute, National Institutes of Health, Department of Health and Human Services, and the Office of Research on Minority Health, Department of Health and Human Services. TRIAL REGISTRATION: clinicaltrials.gov identifier: NCT00006392.

This investigation was supported in part by the following PHS Cooperative Agreement grant numbers awarded by the NCI, DHHS: CA37429, and supported in part by the National Center for Complementary and Alternative Medicine.

References

1. **American Cancer Society.** Cancer Facts and Figures 2008. American Cancer Society, Atlanta, GA, 2008. Available at: http://www.cancer.org/docroot/stt/content/stt_1x_cancer_facts_and_figures_2008.asp (accessed 25 March 2009).
2. **American Cancer Society.** Cancer Facts and Figures for African American 2009–2010. American Cancer Society, Atlanta, GA, 2009. Available at: http://www.cancer.org/docroot/STT/content/STT_1x_Cancer_Facts_Figures_for_African_Americans_2009–2010_09.asp (accessed 25 March 2009).
3. **Ries LAG, Melbert D, Krapcho M,** *et al*. SEER Cancer Statistics Review, 1975–2004, National Cancer Institute. Bethesda, MD Available at: http://seer.cancer.gov/csr/1975_2004/, based on November 2006 SEER data submission, posted to the SEER website, 2007 (accessed 14 December 2007).
4. **Thompson IM, Tangen CM, Klein EA, Lippman SM.** Phase III prostate cancer prevention trials: are the costs justified? *J Clin Oncol* 2005; **23**: 8161–4.
5. **Feigl P, Blumenstein B, Thompson I,** *et al*. Design of the Prostate Cancer Prevention Trial (PCPT). *Control Clin Trials* 1995; **16**: 150–63.
6. **Moinpour CM, Atkinson JO, Thomas SM,** *et al*. Minority recruitment in the prostate cancer prevention trial. *Ann Epidemiol* 2000; **10**: S85–91.
7. **Reynolds T.** PCPT update: enrollment mounts, but minority participation lags. *J Natl Cancer Inst* 1994; **86**: 1500–1.
8. **Coltman CA, Thompson IM, Feigl P.** Prostate Cancer Prevention Trial (PCPT) update. *Eur Urol* 1999; **35**: 544–7.
9. **Swanson GM, Ward AJ.** Recruiting minorities into clinical trials: toward a participant-friendly system. *J Natl Cancer Inst* 1995; **87**: 1747–59.
10. **Woods VD, Montgomery SB, Herring RP.** Recruiting black/African American men for research on prostate cancer prevention. *Cancer* 2004; **100**: 1017–25.
11. **McCaskill-Stevens W, Pinto H, Marcus AC,** *et al*. Recruiting minority cancer patients into cancer clinical trials: a pilot project involving the Eastern Cooperative Oncology Group and the National Medical Association. *J Clin Oncol* 1999; **17**: 1029–39.
12. **Harris Y, Gorelick PB, Samuels P, Bempong I.** Why African Americans may not be participating in clinical trials. *J Natl Med Assoc* 1996; **88**: 630–4.
13. **Millon-Underwood S, Sanders E, Davis M.** Determinants of participation in state-of-the-art cancer prevention, early detection screening, and treatment trials among African-Americans. *Cancer Nurs* 1993; **16**: 25–33.
14. **Green BL, Partridge EE, Fouad MN,** *et al*. African-American attitudes regarding cancer clinical trials and research studies: results from focus group methodology. *Ethn Dis* 2000; **10**: 76–86.
15. **Probstfield JL, Wittes JT, Hunninghake DB.** Recruitment in NHLBI population-based studies and clinical trials: data analysis and survey results. *Control Clin Trials* 1987; **8**: 141S–9S.
16. **Paskett ED, DeGraffinreid CR, Tatum CM, Margitic SE.** The recruitment of African-Americans to cancer prevention and control studies. *Prev Med* 1996; **25**: 547–53.
17. **Probstfield JL.** The clinical trial pre-randomization compliance (adherence) screen. In: Cramer JA, Spilker B. (eds). *Patient Compliance in Medical Practice and Clinical Trials*. Raven Press, New York, 1991, pp. 323–34.
18. **Lippman SM, Goodman PJ, Klein EA,** *et al*. Designing the Selenium and Vitamin E Cancer Prevention Trial (SELECT). *J Natl Cancer Inst* 2005; **97**: 94–102.
19. **Oddone EZ, Olsen MK, Lindquist JH,** *et al*. Enrollment in clinical trials according to patient's race: experience from the VA Cooperative Studies Program (1975–2000). *Control Clin Trials* 2004; **25**: 378–87.
20. **Cook ED, Moody-Thomas S, Anderson KB,** *et al*. Minority recruitment to the Selenium and Vitamin E Cancer Prevention Trial (SELECT). *Clin Trials* 2005; **2**: 436–42.

Erectile Dysfunction Drug Receipt, Risky Sexual Behavior and Sexually Transmitted Diseases in HIV-infected and HIV-uninfected Men

Robert L. Cook, MD, MPH[1], Kathleen A. McGinnis, MS[2], Jeffrey H. Samet, MD[3], David A. Fiellin, MD[4], Maria C. Rodriquez-Barradas, MD[5], Kevin L. Kraemer, MD, MSc[6], Cynthia L. Gibert, MD, MSc[7], R. Scott Braithwaite, MD, MSc, FACP[8], Joseph L. Goulet, PhD, MS[4,9], Kristin Mattocks, PhD, MPH[9], Stephen Crystal, PhD[10], Adam J. Gordon, MD, MPH[2,6], Krisann K. Oursler, MD, ScM[11], and Amy C. Justice, MD, PhD[4,9]

[1]Departments of Epidemiology and Biostatistics and Medicine, University of Florida, Gainesville, FL, USA; [2]Pittsburgh VA Healthcare System, Pittsburgh, PA, USA; [3]Boston University, Boston, MA, USA; [4]Yale University, New Haven, CT, USA; [5]Michael E Debakey VA Medical Center and Baylor College of Medicine, Houston, TX, USA; [6]University of Pittsburgh, Pittsburgh, PA, USA; [7]Washington DC VAMC and George Washington University, Washington, DC, USA; [8]New York University School of Medicine, New York, NY, USA; [9]VA Connecticut Healthcare System, West Haven, CT, USA; [10]Rutgers University, New Brunswick, NJ, USA; [11]Baltimore VA and University of Maryland, Baltimore, MD, USA.

BACKGROUND: Health care providers may be concerned that prescribing erectile dysfunction drugs (EDD) will contribute to risky sexual behavior.

OBJECTIVES: To identify characteristics of men who received EDD prescriptions, determine whether EDD receipt is associated with risky sexual behavior and sexually transmitted diseases (STDs), and determine whether these relationships vary for certain sub-groups.

DESIGN: Cross-sectional study.

PARTICIPANTS: Two thousand seven hundred and eighty-seven sexually-active, HIV-infected and HIV-uninfected men recruited from eight Veterans Health Affairs outpatient clinics. Data were obtained from participant surveys, electronic medical records, and administrative pharmacy data.

MEASURES: EDD receipt was defined as two or more prescriptions for an EDD, risky sex as having unprotected sex with a partner of serodiscordant or unknown HIV status, and STDs, according to self-report.

RESULTS: Overall, 28% of men received EDD in the previous year. Eleven percent of men reported unprotected sex with a serodiscordant/unknown partner in the past year (HIV-infected 15%, HIV-uninfected 6%, P< 0.001). Compared to men who did not receive EDD, men who received EDD were equally likely to report risky sexual behavior (11% vs. 10%, p=0.9) and STDs (7% vs 7%, p=0.7). In multivariate analyses, EDD receipt was not significantly associated with risky sexual behavior or STDs in the entire sample or in subgroups of substance users or men who had sex with men.

CONCLUSION: EDD receipt was common but not associated with risky sexual behavior or STDs in this sample of HIV-infected and uninfected men. However, risky sexual behaviors persist in a minority of HIV-infected men, indicating ongoing need for prevention interventions.

KEY WORDS: HIV infection; risky sexual behavior; STDs; men; phosphodiesterase inhibitors.
J Gen Intern Med 25(2):115–21
DOI: 10.1007/s11606-009-1164-9

Received May 5, 2009
Revised August 21, 2009
Accepted October 15, 2009
Published online November 18, 2009

INTRODUCTION

Phosphodiesterase-5 enzyme inhibitors (sildenafil citrate, tadalafil, and vardenafil HCL) are approved pharmacotherapies to treat erectile dysfunction in men.[1] These erectile dysfunction drugs (EDDs) are commonly used in the United States,[2] marketed broadly, often requested by patients, and associated with improved quality of life.[3] EDDs have also been linked to high-risk sexual behavior in some groups of men at increased risk for HIV transmission, in particular men who have sex with men (MSM), men who use recreational or illicit drugs, and HIV-infected men.[4–16] Because of their association with risky behavior, some have argued that EDD medications should be classified as controlled substances.[5,17]

The source for obtaining EDD is an important issue to consider. Men may obtain EDD via prescription from a health care provider or from other sources, such as the Internet, friends, or the black market.[5,9,10,12] Health care providers must consider whether a prescription of an EDD could have adverse public health effects. Nearly all of the existing literature regarding EDD and risky sexual behavior has come from non-clinical samples of high-risk populations, in which prescribed EDD was rarely differentiated from EDD obtained from other sources. Thus, it is less clear whether prescribing EDD for erectile dysfunction in the context of routine health care is associated with risky sexual behavior or sexually transmitted disease (STD) transmission.

Health care providers caring for HIV-infected men must also consider whether an EDD prescription could facilitate sexual encounters that result in additional HIV transmission. The number of older, HIV-infected men has increased significantly in recent years, largely due to improved treatment but also due to new infections in this age group.[18] As the HIV-infected population ages, the demand for EDDs may increase in this group due to additional chronic diseases that are associated with erectile dysfunction. Thus, it is important to determine whether EDD obtained through routine healthcare is associated with risky sexual behavior in HIV-infected and HIV-uninfected men.

This study's objectives were to identify characteristics of men who received prescriptions for EDD though a network of Veterans Health Affairs (VHA) clinics throughout the United States, to determine whether EDD receipt was associated with risky sexual behavior and sexually transmitted diseases (STDs), and to determine whether these relationships varied based on HIV status, substance abuse, or having sex with men.

METHODS

The Veterans Aging Cohort Study (VACS) is an ongoing prospective cohort study involving HIV-infected and HIV-uninfected veterans receiving care at VHA clinics throughout the United States. The overarching aim of VACS is to study the role of alcohol consumption and comorbid medical and psychiatric disease on clinical outcomes in HIV infection.[19] VACS participants were recruited from infectious disease (HIV-infected) and general medicine (HIV-uninfected) clinics at eight sites (Atlanta, GA; Baltimore, MD; Bronx, NY, Houston, TX; Los Angeles, CA; New York City, NY; Pittsburgh, PA and Washington, DC).[19] Overall, 58% of HIV-infected men at the eight sites were enrolled, with only 9% of those approached refusing to participate.[19] HIV uninfected controls were targeted to match the demographics of the HIV-infected participants on 5-year age blocks, race and gender. Subjects completed a comprehensive baseline survey at enrollment and then at one year follow-up intervals. Further descriptions of the VACS sample and methodology are available online (www.vacohort.org).

The study sample for these analyses includes the subset of male VACS participants who completed a follow-up survey between September 2005 and January 2007 and who reported any sexual activity in the past year (n=2,787). Data were obtained from three linked sources: participant surveys, electronic medical records, and pharmacy data that are collected nationally through the Pharmacy Benefits Management (PBM) program (Hines, IL). The PBM program includes all outpatient prescriptions funded through the VHA healthcare system and are likely to be representative of the use of prescribed EDD in this population.

Measures. *EDD use* was defined as two or more prescriptions for sildenafil citrate, tadalafil, or vardenafil HCL, documented in the PBM database in the year prior to and up to the follow-up survey date. Thus, men who had received only one EDD prescription were not considered to be EDD users, because such men may have never used the medication, or simply tried it once. Although men were not asked about non-prescribed use of EDD at this assessment, a previous assessment in the same cohort demonstrated substantial agreement between self-reported EDD use and the PMB database, with fewer than 5% of men reporting EDD use exclusively from non-VA sources (unpublished data).

Risky sexual behavior was defined as "unprotected sex with a partner of serodiscordant or unknown HIV status". Specifically, men were asked, "During the past 12 months, did you ever, even once, have unprotected vaginal or anal sex (sex without a condom) with any of the following types of partners? Any partner who was HIV positive; Any partner who was HIV negative; Any partner whose HIV status was unknown." The following were the response options: Yes (unprotected sex at least once); or No (always used a condom). Those who reported unprotected sex with anyone of serodiscordant or unknown HIV status were classified as having risky sexual behavior.

STDs were identified by self report. Participants were asked, "In the past 12 months, have you been diagnosed with any of the following sexually transmitted diseases?" Persons were classified as having had a self-reported STD if they reported having genital warts, chlamydia, gonorrhea, syphilis, trichomonas, chancroid, or herpes.

Participant demographic characteristics, including gender, race/ethnicity, and marital status/living with partner were measured by self-report on the VACS baseline survey. Measures of alcohol and drug use and depression were included on the VACS follow-up surveys. Hazardous drinking (drinking associated with possible harm) was defined as a score of eight or more on the Alcohol Use Disorders Identification Test (AUDIT).[20] Participants were asked to report the frequency of use of marijuana, cocaine, stimulants, and heroin; current use for these drugs was defined as "at least monthly." Non-prescribed use of pain medications, defined as use of pain medications that were not prescribed, was based on self-reported use in the past year from a list of 20 specific narcotic pain medications. Depression, measured by the Patient Health Questionnaire (PHQ-9) at each follow-up assessment, was classified as present if the PHQ-9 score was 10 or more.[21] Comorbid medical conditions, including diagnoses of hypertension, diabetes, and coronary artery disease (CAD), were determined using International Classification of Diseases, Ninth Revision (ICD-9) codes from the electronic medical record. Persons were considered to have a comorbid diagnosis if at least one inpatient or two outpatient ICD-9 code diagnoses were recorded between one-year prior to and six months after the survey date. Persons were classified as having sex with men if they reported having "sex with males" or "sex with males and females" in the past year. Further details on variables and surveys can be found at www.vacohort.org.

Analysis

Demographic and behavioral characteristics were described and compared by HIV status using chi-square tests and t-tests, as appropriate. We used chi-square tests to determine whether EDD receipt varied by demographic and descriptive variables and to determine the bivariate associations of EDD receipt to risky sex and STDs.

Multivariable logistic regression models were used to identify independent factors associated with EDD receipt, risky sexual behavior, and STDs. Logistic regression models predicting EDD receipt were adjusted for demographic variables (age, race/ethnicity, married/living with partner, sex with

males in past 12 months), alcohol and drug use, and comorbidities identified as potential confounders (depression, diabetes, hypertension, and coronary artery disease). The logistic regression models predicting risky sexual behavior and STD included the same variables plus EDD receipt. There was significant collinearity between the sex with males and HIV status variables; therefore, we ran multivariate models stratified by HIV status. For each set of models, we first ran the multivariate logistic regression models including the variables described above. We then reran the models excluding variables with p>0.2 in both HIV-infected and uninfected models, and we present these models in the results. We also examined the relationship of EDD to risky sexual behavior or STDs in the following subgroups: those with hazardous drinking, cocaine use, and sex with men. All analyses were conducted using Stata 10.0 (College Station, TX).

Multiple imputation was used to address missing data.[22–24] For each of the following covariates, 1% or less were missing a response: married or living with partner, had sex with males, hazardous alcohol use, and depression. Pain medication use information was missing for 2%, and marijuana and cocaine use information was missing for 4%. Multiple imputation was conducted using the Stata v10.0 (Stata Corporation, College Station, Texas) ice command.[22–24] The imputation model included EDD receipt, risky sexual behavior, and all covariates included in the initial models (listed above), undertook ten switching procedures, and generated five datasets. The Stata v10.0 mim command was used to combine the results of the analyses from the multiply imputed data sets.[23] Analyses were also conducted on the complete case dataset (using categories to define missing data). Results from analyses of imputed data did not differ substantively from those using complete cases; therefore, the results from the imputation models are included here.

RESULTS

The study sample consists of 2,787 sexually-active men, of whom 1,469 (53%) were HIV-infected. Demographic and behavioral characteristics in both HIV-infected and HIV-uninfected men are shown in Table 1. Over 60% of the men were over age 50, over two-thirds were black, and one-third were married or living with a long-term partner. Many had comorbid health conditions, and substance use behavior was common; 18% reported hazardous alcohol consumption, and over 10%

Table 1. Demographics and Main Outcomes Among 2,787 Sexually Active Men Participating in VACS, 2005–2007, Overall and by HIV Serostatus

	ALL	HIV-infected	HIV-uninfected	P-Value
N	2,787	1,469	1,318	
Mean age (SD)	52.0 (8.9)	51 (8.5)	53 (9.1)	<0.001
Age in years, N (%)				
<40	215 (8)	133 (9)	82 (6)	<0.001
40-49	874 (31)	494 (34)	380 (29)	
50+	1,698 (61)	842 (57)	856 (65)	
Race/Ethnicity, N (%)				0.3
White	513 (18)	254 (17)	259 (20)	
Black	1926 (69)	1036 (71)	890 (68)	
Hispanic	241 (9)	126 (9)	115 (9)	
Other/Unknown	107 (4)	53 (4)	54 (4)	
Married or living with partner, N (%)[a]	915 (33)	376 (26)	539 (41)	<0.001
Sex with men in past year, N (%)[a]	636 (23)	611 (42)	25 (2)	<0.001
Alcohol and drug use, N (%)[a]				
Hazardous alcohol consumption[b]	500 (18)	240 (17)	260 (20)	0.019
Non-prescribed pain medication[c]	374 (14)	216 (15)	158 (12)	0.04
Marijuana (monthly)	358 (13)	234 (17)	124 (10)	<0.001
Cocaine (monthly)	312 (12)	176 (13)	136 (11)	0.2
Stimulants (monthly)	40 (2)	34 (2)	6 (0.5)	<0.001
Heroin (monthly)	91 (3)	44 (3)	47 (4)	0.4
Comorbid health conditions, N (%)				
Depression[d]	566 (20)	257 (18)	309 (24)	<0.001
Hypertension[e]	1154 (41)	409 (28)	745 (57)	<0.001
Diabetes[e]	468 (17)	172 (12)	296 (22)	<0.001
Coronary artery disease[e]	196 (7)	58 (4)	138 (10)	<0.001
Peripheral vascular disease[e]	72 (3)	22 (2)	50 (4)	<0.001
EDD receipt in past year, N (%)	788 (28)	380 (26)	408 (31)	0.03
Risky sexual behavior (past year), N (%)[f]	284 (11)	214 (15)	70 (6)	<0.001
Sexually transmitted diseases (past year), N (%)[a]	200 (7)	145 (10)	55 (4)	<0.001
Genital herpes	92 (3)	57 (4)	35 (3)	0.07
Genital warts	47 (2)	31 (2)	16 (1)	0.07
Chlamydia	55 (2)	29 (2)	26 (2)	0.99
Gonorrhea	60 (2)	37 (3)	23 (2)	0.16
Syphilis	66 (2)	45 (3)	21 (2)	0.01
Chancroid	27 (1)	10 (1)	17 (1)	0.1
Trichomonas	32 (1)	13 (1)	19 (1)	0.17

P-value: comparison of HIV-infected vs. HIV-uninfected [a]Based on self-report; [b]Hazardous Alcohol Consumption: Alcohol Use Disorders Identification Test (AUDIT) score of ≥8; [c]Non-prescribed pain medications defined as any use in past year from list of 20 narcotic pain medications; [d]Score of 10 or more on the Patient Health Questionnaire (PHQ-9) (21); [e]Based on ICD-9 diagnosis; [f]defined as unprotected intercourse with a partner of serodiscordant or unknown HIV serostatus

reported marijuana, cocaine, or the non-prescribed use of pain medications. Compared to HIV-uninfected men, HIV-infected men were more likely to have reported having had sex with men, and the use of marijuana, non-prescribed pain medications, and stimulants (Table 1). HIV-uninfected men were more likely to be married or living with a partner, have hazardous alcohol consumption, and have other medical comorbidities.

Overall, 28% of these sexually active men received two or more prescriptions for EDD in the previous year (Table 1). HIV-infected men were slightly less likely to be prescribed EDD (26% vs. 31%, p=0.003). Regarding risky sexual behavior, 11% reported unprotected sex with a non-main partner in the past year. HIV-infected men were significantly more likely than HIV-uninfected men to engage in risky sexual behavior (15% vs. 6%, p<0.001). STDs were reported by 7% of the sample. HIV-infected men were significantly more likely to report a STD diagnosis in the previous year (10% vs. 4% p<0.001).

Individual Characteristics Associated with EDD Prescription

One-third of men age 50 years and over received two or more prescriptions for EDD in the previous year. Table 2 shows the relationship of demographic, behavioral, and clinical characteristics with EDD use overall, and among HIV-infected and HIV-uninfected men. Overall, EDD receipt was more common in older men, nonwhite men, non-MSM men, those with non-prescribed use of pain medications, and those with depression, hypertension, or diabetes (Table 2). Among HIV-infected men, EDD receipt was more common in older men, men who don't have sex with men, those with non-prescribed use of pain medications, and those with diabetes (Table 2). Among HIV-uninfected men, EDD receipt was more common in older men, nonwhite men, those not married or living with partner, and those with non-prescribed use of pain medications, cocaine use, depression, or hypertension (Table 2).

In both the HIV-infected and HIV-uninfected multivariate models, increasing age and use of unprescribed pain medications were associated with an increased likelihood of EDD receipt (Table 3). In the HIV-infected model, having sex with men was associated with a decreased likelihood of EDD receipt. In the HIV-uninfected model, non-white race/ethnicity and depression were associated with an increased likelihood of EDD receipt, while married/ living with partner was associated with a decreased likelihood of EDD receipt (Table 3).

Association of EDD Receipt with Risky Sexual Behavior and STDs

The percent of reported risky sexual behavior was similar between those who did and did not receive EDD (11% vs. 10%, p=0.6). This was true when the analysis was restricted to the HIV-infected men (16% vs. 15%, p=0.4) or the HIV-uninfected men (6% vs. 5%, p=0.6).

Table 2. Factors Associated with EDD Receipt Among Sexually Active Men Participating in VACS, 2005–2007, Overall and in HIV-infected and HIV-uninfected Men

Patient characteristic		Proportion of persons who received EDD in past year, N (%)		
		ALL (n=2,787)	HIV-infected (n=1,469)	HIV-uninfected (n=1,318)
Age in years	<40	14 (7)	9 (7)	5 (6)
	40-49	210 (24)	114 (23)	96 (25)
	50+	564 (33)[g]	257 (31)[g]	307 (36)[g]
Race/ethncity	White	108 (21)	61 (24)	47 (18)
	Black	583 (30)	270 (26)	313 (35)
	Hispanic	65 (27)	34 (27)	31 (27)
	Other/Unknown	32 (30)[g]	15 (28)	17 (31)[g]
Married or living with partner[a]	Yes	242 (26)	106 (28)	136 (25)
	No	538 (29)	267 (25)	271 (35)[g]
Sex with men in past year[a]	Yes	116 (18)	112 (18)	4 (16)
	No	668 (31)[g]	265 (32)[g]	403 (31)
Hazardous alcohol consumption[b]	Yes	144 (29)	61 (25)	83 (32)
	No	631 (28)	313 (26)	318 (31)
Non-prescribed pain medication[c]	Yes	148 (40)	74 (34)	74 (47)
	No	616 (26)[g]	297 (24)[g]	319 (28)[g]
Marijuana use[a]	Yes	93 (26)	53 (23)	40 (32)
	No	654 (28)	311 (26)	343 (30)
Cocaine use[a]	Yes	96 (31)	43 (24)	53 (39)
	No	650 (28)	320 (26)	330 (30)[f]
Stimulants use[a]	Yes	7 (18)	5 (15)	2 (33)
	No	727 (28)	353 (26)	374 (30)
Depression[d]	Yes	189 (33)	77 (30)	112 (36)
	No	591 (27)[g]	299 (25)	292 (29)[f]
Hypertension[e]	Yes	379 (33)	126 (31)	253 (34)
	No	409 (25)[g]	254 (24)[g]	155 (27)[g]
Diabetes[e]	Yes	162 (35)	58 (34)	104 (35)
	No	626 (27)[f]	322 (25)[f]	304 (30)
Coronary artery disease[e]	Yes	51 (26)	16 (28)	35 (25)
	No	737 (28)	364 (26)	373 (32)

[a]Based on self-report; [b]Hazardous Alcohol Consumption: Alcohol Use Disorders Identification Test (AUDIT) score of ≥8; [c]Non-prescribed pain medications defined as any use in past year from list of 20 narcotic pain medications; [d]Score of 10 or more on the Patient Health Questionnaire (PHQ-9) (21); [e]Based on ICD-9 diagnosis [f]p<0.05; [g]p<0.01; p-values from chi-square tests for differences in percentage of persons who received EDD in each of the patient characteristic categories

Table 3. Factors Associated with EDD Receipt Among Sexually Active Men Participating in VACS, 2005–2007: Multivariate Analysis

	HIV-infected (n=1,469)			HIV-uninfected (n=1,318)		
	OR	95% CI	P	OR	95% CI	P
Age						
<40 years	–			–		
40-49	3.4	1.6-6.9	0.001	4.3	1.7-11.0	0.003
50+	4.3	2.1-8.7	<0.001	7.8	3.1-19.8	<0.001
Non-white (ref=white)	0.8	0.6-1.2	0.3	2.1	1.5-3.0	<0.001
Married/living with partner[a]	1.1	0.8-1.4	0.6	0.6	0.5-0.8	0.001
Sex with men in past year[a] (ref=no sex with men in past year)	0.6	0.4-0.8	<0.001	0.5	0.2-1.4	0.2
Non-prescribed pain medication in past year[b] (ref=none)	1.4	1.0-1.9	0.03	2.0	1.4-2.9	<0.001
Depression[c] (ref=none)	1.2	0.9-1.7	0.2	1.4	1.0-1.8	0.02
Hypertension[d] (ref=none)	1.2	0.9-1.6	0.1	1.3	1.0-1.7	0.07
Coronary artery disease[d] (ref=none)	0.8	0.4-1.5	0.5	0.7	0.4-1.0	0.06

[a]Based on self-report; [b]Non-prescribed pain medications defined as any use in past year from list of 20 narcotic pain medications; [c]Score of 10 or more on the Patient Health Questionnaire (PHQ-9) (21); [d]Based on ICD-9 diagnosis

In both the HIV-infected and HIV-uninfected models, being married/living with partner were associated with decreased likelihood of risky sex, whereas sex with males, hazardous alcohol consumption, and cocaine use were associated with increased likelihood of risky sex (Table 4). In the HIV-infected model only, unprescribed pain medication and marijuana use was associated with an increased likelihood of risky sex. For both HIV-infected and uninfected models, EDD receipt was not statistically significantly associated with risky sex (Table 4).

Overall, the percentage of reported STDs was similar between those who did not receive EDD (7% vs. 7%, p=0.7). This finding was similar among HIV-infected men (9% vs. 10%, p=0.5); although among HIV-uninfected men, those who received EDD were more likely to report an STD in bivariate analyses (6% vs. 3%, p=0.02). For both HIV-infected and uninfected models, EDD receipt was not statistically significantly associated with STDs (Table 5).In both the HIV-infected and HIV-uninfected multivariate models of STDs, unprescribed pain medication was associated with an increased likelihood of STDs. In the HIV-infected model, younger age and sex with men was associated with an increased likelihood of STDs.

There were no statistically significant relationships between EDD receipt and risky sexual behaviors or STDs in models limited to those with hazardous drinking, cocaine use, or men who reported having sex with men (data not shown).

DISCUSSION

In this sample of over 2,500 men attending VHA outpatient clinics, EDD receipt was not associated with risky sexual behavior or STDs, overall or within subgroups of HIV-infected men, substance users, or MSM. Thus, for men who obtain EDD via prescription from a healthcare provider, EDD appears to be prescribed responsibly and used responsibly. These findings differ from the majority of previous studies on this topic. One of the most plausible reasons for the varying conclusions is that the focus of our study is on men who received EDD as part of clinical practice, whereas nearly all of the previous reports linking EDD to risky sexual behavior were conducted in non-clinical samples in high risk groups, including MSM [4–12,16,25], substance abusers[11,26], and HIV-infected men outside of clinical settings.[13–15] In nearly all of these studies, the association of EDD and risky sexual behavior was consistently the strongest in men who also used stimulant drugs such as methamphetamines, ecstasy, or gamma-hydroxybutyrate (GHB), which were rare in our study sample.

The source of EDD may also influence its relationship with risky sexual behavior. In samples of younger MSMs, 40% or more report obtaining EDD without a prescription (e.g. via the internet or off the street).[5,9,10,12] Men obtaining EDD without a prescription appear to report higher-risk behaviors than those

Table 4. Factors Associated with Risky Sexual Behavior in Sexually Active Men Participating in VACS, 2005–2007: Multivariate Analysis

	HIV-infected (n=1,469)			HIV-uninfected (n=1,318)		
	OR	95% CI	P	OR	95% CI	P
EDD receipt	1.2	0.9-1.8	0.2	1.0	0.6-1.7	0.9
Non-white (ref=White)	1.0	0.6-1.5	0.9	1.7	0.8-3.7	0.2
Married/living w/ partner[a]	0.6	0.4-0.9	0.03	0.4	0.2-0.7	0.004
Sex with men in past year[a] (ref=no sex with men in past year)	2.0	1.4-2.8	<0.001	4.0	1.3-11.7	0.01
Hazardous alcohol consumption[b] (ref=none)	1.7	1.1-2.5	0.008	2.2	1.3-3.8	0.004
Non-prescribed pain medication[c] (ref=none)	1.6	1.1-2.4	0.02	0.5	0.2-1.1	0.09
Marijuana use[a] (ref=none)	1.5	1.0-2.1	0.04	1.8	0.9-3.6	0.08
Cocaine use[a] (ref=none)	1.8	1.2-2.7	0.01	2.4	1.3-4.5	0.005
Depression[d] (ref=none)	1.1	0.7-1.6	0.7	1.5	0.9-2.5	0.2

Risky sexual behavior defined as unprotected intercourse with a partner of serodiscordant or unknown HIV serostatus[a]Based on self-report; [b]Hazardous Alcohol Consumption: Alcohol Use Disorders Identification Test (AUDIT) score of ≥8; [c]Non-prescribed pain medications defined as any use in past year from list of 20 narcotic pain medications; [d]Score of 10 or more on the Patient Health Questionnaire (PHQ-9) (21)

Table 5. Factors Associated with Self-Reported Sexually Transmitted Diseases (STDs) in Sexually Active Men Participating in VACS, 2005–2007: Multivariate Analysis

	HIV-infected (n=1,469)			HIV-uninfected (n=1,318)		
	OR	95% CI	P	OR	95% CI	P
EDD receipt	0.99	0.6-1.5	0.9	1.7	0.9-2.9	0.09
Age						
<40	–			–		
40-49	0.5	0.3-0.8	0.006	0.8	0.2-3.0	0.8
50+	0.6	0.3-1.0	0.038	1.2	0.3-4.3	0.8
Non-White (ref=White)	1.5	0.9-2.6	0.1	1.8	0.7-4.4	0.2
Sex with men in past year[a] (ref=no sex w/ men in past year)	1.5	1.0-2.3	0.04	2.1	0.4-10.0	0.4
Hazardous alcohol consumption[b] (ref=none)	1.5	0.9-2.3	0.1	1.7	0.9-3.1	0.08
Non-prescribed pain medication[c] (ref=none)	1.9	1.2-3.0	0.004	2.0	1.0-3.8	0.04
Marijuana use[a] (ref=none)	1.2	0.8-1.9	0.4	1.8	0.9-3.8	0.1
Depression[d] (ref=none)	1.4	0.9-2.2	0.1	1.4	0.8-2.5	0.3
Hypertension[e] (ref=none)	0.7	0.4-1.1	0.1	0.7	0.4-1.2	0.2
Diabetes[e] (ref=none)	0.5	0.3-1.1	0.1	0.7	0.3-1.5	0.3

[a]Based on self-report; [b]Hazardous Alcohol Consumption: Alcohol Use Disorders Identification Test (AUDIT) score of ≥8; [c]Non-prescribed pain medications defined as any use in past year from list of 20 narcotic pain medications; [d]Score of 10 or more on the Patient Health Questionnaire (PHQ-9) (21); [e]Based on ICD-9 diagnosis

obtaining it by prescription.[9] A study of EDD use in heterosexual drug users found that many men used the medications to "enhance sexual experience," rather than to treat erectile dysfunction, but 30% of that sample had obtained these medications without a prescription.[11] The current study only considered EDD that had been obtained by prescription, a source that is most relevant and under the control of the prescribing clinician.

In this study, the proportion of sexually active men who received prescriptions for EDD use was fairly high; one-third of men over age 50 received two or more prescriptions in the previous year, with receipt being equally likely among men aged 50 to 60 as in those aged 60 and above. EDD receipt was less common among MSM in this sample, which could reflect either less erectile dysfunction, or decreased use of physicians as a source of EDDs**,** or a reluctance by physicians to prescribe EDD to these patients. The association of EDD receipt with non-prescribed use of pain medications had not been reported previously. Possible explanations include an association of chronic pain with erectile dysfunction, side effects of the specific medications, or other characteristics of men that are more likely to use pain medications without a prescription. Similarly, the finding linking depression with EDD receipt could reflect either depression as a cause of erectile dysfunction, erectile dysfunction as a cause of depression, or side effects of medications used to treat depression.[27,28] Other studies have also found an association of depression with EDD use in HIV-infected men.[29]

The proportion of sexually active men who engaged in risky sex or reported STDs was fairly high, especially among HIV-infected men. This is consistent with a previous analysis in VACS in which men were more likely to engage in risky sex if they were younger, reported hazardous alcohol or drug use, were not married, or were MSM.[30] Persons with these risk factors may benefit from additional attention regarding HIV/STD prevention counseling regardless of their EDD use; an opportune time to address sexual health risks would be when an EDD is being prescribed.

Several potential study limitations should be noted. As in any cross-sectional analysis, the cause-and-effect relationship between EDD receipt and risky sexual behaviors can be difficult to assess, especially when the associations are compared at general levels rather than event-specific analyses. Men attending VHA clinics may have sociodemographic or behavioral characteristics that are different from men recruited from non-VHA clinical settings, although veterans have not been the focus of prior research on this topic. Our definition of EDD receipt does not include EDD that was obtained from a non-VHA pharmacy, provided as a sample medication, or obtained without a prescription. However, the focus of this analysis is on EDD obtained from a healthcare provider. It is also possible that providers systematically declined to provide EDD to men that they knew or suspected were engaging in risky sexual behavior, but this was not measured in the current study. Finally, measures of risky sexual behavior and self-reported STDs have limitations and are likely to underestimate the true prevalence of these behaviors and infections. However, we have no reason to suspect that reporting of these conditions would vary according to EDD receipt.

In conclusion, the findings from this study provide some reassurance to healthcare providers who prescribe EDD, including those who provide care to HIV-infected men. Although it is clear that some men who receive EDD engage in risky sexual behavior, we found that HIV-infected and HIV-uninfected men who received EDD were no more likely to engage in risky sexual behavior or to report new STDs than men who did not receive EDD. Physicians should continue to counsel their patients about HIV/STD prevention at opportune moments such as discussion of EDD, general checkups, or when identifying other risk factors associated with risky sexual behavior.

Acknowledgements: *These data were presented, in part, at the 31st Annual SGIM Meeting, Pittsburgh PA, April 2008.*

Funding support: *VACS is supported by National Institute on Alcohol and Alcohol Abuse (3U01 AA 13566), National Institute of Aging (K23 AG00826), Robert Wood Johnson Generalist Faculty Scholar Award, an Inter-agency Agreement between National Institute on Aging, National Institute of Mental Health and the*

Veterans Health Administration, and the Veterans Health Administration Office of Research and Development and Public Health Strategic Health Care Group. Dr. Crystal's work is supported in part by AHRQ grant U18-HS016997, Center for Education and Research on Mental Health Therapeutics. The funding agencies had no specific role or restrictions regarding the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript. Dr. Cook 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.

Conflict of Interest: *None disclosed.*

Disclaimer: *The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of the Department of Veterans Affairs.*

Corresponding Author: *Robert L. Cook, MD, MPH; Departments of Epidemiology and Biostatistics and Medicine, University of Florida, PO Box 100231, Gainesville, FL 32610, USA (e-mail: cookrl@phhp.ufl.edu).*

REFERENCES

1. **Lue TF.** Erectile dysfunction. N Engl J Med. 2000;342(24):1802–13.
2. **Lindau ST, Schumm LP, Laumann EO, Levinson W, O'Muircheartaigh CA, Waite LJ.** A study of sexuality and health among older adults in the United States. N Engl J Med. 2007;357(8):762–74.
3. **Smith KJ, Roberts MS.** The cost-effectiveness of sildenafil. Ann Intern Med. 2000;132(12):933–7.
4. **Spindler HH, Scheer S, Chen SY, et al.** Viagra, methamphetamine, and HIV risk: results from a probability sample of MSM, San Francisco. Sex Transm Dis. 2007;34(8):586–91.
5. **Swearingen SG, Klausner JD.** Sildenafil use, sexual risk behavior, and risk for sexually transmitted diseases, including HIV infection. Am J Med. 2005;118(6):571–7.
6. **Crosby R, Diclemente RJ.** Use of recreational Viagra among men having sex with men. Sex Transm Infect. 2004;80(6):466–8.
7. **Paul JP, Pollack L, Osmond D, Catania JA.** Viagra (sildenafil) use in a population-based sample of U.S. men who have sex with men. Sex Transm Dis. 2005;32(9):531–3.
8. **Mansergh G, Shouse RL, Marks G, et al.** Methamphetamine and sildenafil (Viagra) use are linked to unprotected receptive and insertive anal sex, respectively, in a sample of men who have sex with men. Sex Transm Infect. 2006;82(2):131–4.
9. **Marks G, Richardson JL, Milam J, Bolan R, Stoyanoff S, McCutchan A.** Use of erectile dysfunction medication and unsafe sex among HIV+ men who have sex with men in care. Int J STD AIDS. 2005;16(3):271–2.
10. **Pantalone DW, Bimbi DS, Parsons JT.** Motivations for the recreational use of erectile enhancing medications in urban gay and bisexual men. Sex Transm Infect. 2008;84(6):458–62.
11. **Horvath KJ, Calsyn DA, Terry C, Cotton A.** Erectile dysfunction medication use among men seeking substance abuse treatment. J Addict Dis. 2007;26(4):7–13.
12. **Fisher DG, Reynolds GL, Ware MR, Napper LE.** Methamphetamine and Viagra use: Relationship to sexual risk behaviors. Arch Sex Behav. 2009; epub Mar 28 2009.
13. **Purcell DW, Wolitski RJ, Hoff CC, Parsons JT, Woods WJ, Halkitis PN.** Predictors of the use of viagra, testosterone, and antidepressants among HIV-seropositive gay and bisexual men. AIDS. 2005;19(Suppl 1):S57–66.
14. **Cooperman NA, Arnsten JH, Klein RS.** Current sexual activity and risky sexual behavior in older men with or at risk for HIV infection. AIDS Educ Prev. 2007;19(4):321–33.
15. **Cachay E, Mar-Tang M, Mathews WC.** Screening for potentially transmitting sexual risk behaviors, urethral sexually transmitted infection, and sildenafil use among males entering care for HIV infection. AIDS Patient Care Stds. 2004;18(6):349–54.
16. **Ostrow DG, Plankey MW, Cox C, et al.** Specific sex drug combinations contribute to the majority of recent HIV seroconversions among MSM in the MACS. J Acquir Immune Defic Syndr. 2009;51(3):349–55.
17. **Hulgan T.** Klausner JD: Phosphodiesterase type-5 inhibitors and the reemerging HIV epidemic. JAMA. 2008;299(12):1426.
18. Centers for Disease Control and Prevention. HIV/AIDS among Persons Aged 50 and Older. 2008. Available at: http://www.cdc.gov/hiv/topics/over50/resources/factsheets/over50.htm., accessed October 9, 2009.
19. **Justice AC, Dombrowski E, Conigliaro J, et al.** Veterans Aging Cohort Study (VACS): Overview and description. Med Care. 2006;44(8 Suppl 2): S13–24.
20. **Babor T, Ramon Dela Fuente J, Saunders J, Grant M.** The alcohol use dsorders identification test: guidelines for use in primary health care. Geneva: World Health Organization; 1989.
21. **Kroenke K, Spitzer RL, Williams JB.** The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med. 2001;16(9):606–13.
22. **Sterne JA, White IR, Carlin JB, et al.** Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ. 2009;338:b2393.
23. **Royston P.** Multiple imputation of missing values: further update of ice, with an emphasis on interval censoring. Stata Journal. 2007;7:445–64.
24. **van Buuren S, Boshuizen HC, Knook DL.** Multiple imputation of missing blood pressure covariates in survival analysis. Stat Med. 1999;18(6):681–94.
25. **Drumright LN, Gorbach PM, Little SJ, Strathdee SA.** Associations between substance use, erectile dysfunction medication and recent HIV infection among men who have sex with men. AIDS Behav. 2009;13 (2):328–36.
26. **Fisher DG, Malow R, Rosenberg R, Reynolds GL, Farrell N, Jaffe A.** Recreational Viagra use and sexual risk among drug abusing men. Am J Infect Dis. 2006;2(2):107–14.
27. **Hackett G.** The burden and extent of comorbid conditions in patients with erectile dysfunction. Int J Clin Pract. 2009;63(8):1205–13.
28. **Kennedy SH, Rizvi S.** Sexual dysfunction, depression, and the impact of antidepressants. J Clin Psychopharmacol. 2009;29(2):157–64.
29. **Crum-Cianflone NF, Bavaro M, Hale B, et al.** Erectile dysfunction and hypogonadism among men with HIV. AIDS Patient Care STDS. 2007;21 (1):9–19.
30. **Cook RL, McGinnis KA, Kraemer KL, et al.** Intoxication before intercourse and risky sexual behavior in male veterans with and without human immunodeficiency virus infection. Med Care. 2006;44(8 Suppl 2):S31–6.

Journal of Health Communication, 15:172–182, 2010

ISSN: 1081-0730 print/1087-0415 online
DOI: 10.1080/10810730.2010.499990

Routledge
Taylor & Francis Group

How to Achieve Informed Consent for Research from Spanish-Speaking Individuals with Low Literacy: A Qualitative Report

DHARMA E. CORTÉS

Department of Psychiatry, Cambridge Hospital/Harvard Medical School, Cambridge, Massachusetts, USA

MARI-LYNN DRAINONI

Department of Health Policy and Management, Boston University School of Public Health, Boston, and the Center for Health Quality, Outcomes and Economic Research, ENRM VA Hospital, Bedford, Massachusetts, USA

LORI E. HENAULT AND
MICHAEL K. PAASCHE-ORLOW

Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, Boston, Massachusetts, USA

Investigators have the responsibility to ensure that prospective participants are fully informed about a research protocol prior to consenting to participate, yet many researchers face challenges when obtaining consent, since the majority of the general population has limited or no familiarity with research studies. These challenges are further magnified when obtaining consent from individuals with low literacy levels and who speak languages other than English. In this article we present findings from a qualitative study conducted with Spanish-speaking individuals with low-literacy designed to refine the Agency for Healthcare Research and Quality's Informed Consent and Authorization Toolkit for Minimal Risk Research. Findings from this study indicate that familiarity with providing informed consent and authorization for research or the experience of being a research participant appear to play key roles in an individual's ability to understand the consent and authorization process. While the text of the consent and authorization documents can be simplified using plain language principles, comprehension of several fundamental ideas such as risk and privacy need to be safeguarded with a consent process that confirms comprehension. Recommendations are provided to address the informational needs of individuals with low literacy levels and limited or no experience with research participation.

Investigators have the responsibility to ensure that prospective participants are fully informed about a research protocol prior to consenting to participate. This responsibility is fulfilled by means of an informed consent and the Health Insurance

Address correspondence to Dharma E. Cortés, PhD, Department of Psychiatry, Cambridge Hospital/Harvard Medical School, 1493 Cambridge Street, Cambridge, MA 02139, USA. E-mail: dharma_cortes@hms.harvard.edu

Portability and Accountability Act (HIPAA) authorization process [http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.htm; http://www.hhs.gov/ocr/privacy/]. Barring special circumstances, this process is facilitated and documented by means of an informed consent and HIPAA authorization form. Further, federal regulations mandate that consent for human research participants be presented in a fashion that is 'understandable' by potential participants and that research authorization forms be written in 'plain language' [http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.htm]; [http://www.hhs.gov/ocr/privacy/hipaa/administrative/privacyrule/adminsimpregtext.pdf].

Consequently, many Institutional Review Boards (IRBs) have established readability guidelines. Unfortunately, IRBs rarely meet their own guidelines in the template text that they require investigators to use and most research consent forms present information in a manner that is unnecessarily complex and unreadable by most people (Paasche-Orlow et al., 2003). As such, obtaining consent—using written consent forms—for research participation from individuals with low literacy poses great challenges (Davis et al., 1998).

Research participants have commonly been found to lack basic understanding regarding fundamental aspects of the studies in which they are participating (Flory and Emanuel, 2004; Jenkins et al., 2010; Raich et al., 2001; Sugarman et al., 1999). While most authors have found benefits from designing consent forms that are shorter and easier to read (Coyne et al., 2003; Flory and Emanuel, 2004; Kang et al., 2009; Paris et al., 2010) simply decreasing the reading level will not ensure that all potential participants give consent and authorization that have been substantively informed. For example, easy-to-read consent forms might not be enough for potential research participants who lack familiarity with *providing* informed consent or the activities of participating in research (Sudore et al., 2006).

In addition, making consent forms easier to read does not necessarily address the linguistic challenges that researchers face when obtaining consent from individuals who speak languages other than English and/or have limited English proficiency even when using a consent form that has been translated into the target language without determining whether the translation process rendered an easy-to-understand consent form. To address this situation, most IRBs overseeing research conducted with non-English speaking populations allow investigators either to use a short-form translated generic statement delivered by an interpreter along with a discussion of the research using the English consent form as a guide or require researchers to submit translated versions of the consent forms they intend to use [REF: http://www.hhs.gov/ohrp/humansubjects/guidance/ic-non-e.htm] frequently requiring the application of a back-translation method (i.e., the process of translating a document that has already been translated into another language back to the original language by an independent translator) as a mechanism to assess the translation's accuracy. Although these strategies represent significant improvements from past practices, they may not be enough.

To address many of these concerns, an Informed Consent and HIPAA Authorization Toolkit was developed by Agency for Healthcare Research and Quality (AHRQ) staff and modified with input from experts in health literacy, clinical research, Federal regulations, HIPAA authorization requirements, and ethics. The toolkit was also pre-tested by researchers at Boston University Medical Center (ACTION Network Field Partnerships for Applied Research, Principal Investigator: MP-O) with a diverse group of research participants, researchers, and IRB

professionals. To refine and validate the language of the sample consent documents in this toolkit, we recruited English- and Spanish-speaking research participants with different levels of literacy to participate in focus groups and individual interviews in five cities (Atlanta, GA; Baltimore, MD; Boston, MA; Tucson, AZ; and Waukegan, IL). These materials are not copyright protected (http://www.ahrq.gov/fund/informedconsent/) and have been reviewed to ensure that the guidance provided is consistent with the regulations for obtaining and documenting informed consent for participation in minimal risk research and authorization for use of protected health information as required under HIPAA (see Appendix for example text excerpt). In this article we present findings from the qualitative study conducted with Spanish-speaking individuals with low-literacy designed to assess comprehension of and refine the AHRQ Informed Consent and Authorization Toolkit for Minimal Risk Research.

Methods

To evaluate the comprehension of a Spanish version (translated from English) of an easy-to-read minimal risk consent form and HIPAA authorization among Spanish-speaking Latino/as, we conducted three focus groups and 12 individual interviews. Individuals were recruited at adult basic education (ABE) programs and health clinics. *Teachers* at the ABE programs and *clinical staff* from the health clinics identified Spanish-speakers with low and higher literacy and informed them about the study. If they agreed to participate, they were invited to be part of a focus group or participate in an individual interview with a Spanish-speaking PhD level researcher. A total of 37 individuals were recruited. Twelve participated in twelve qualitative individual interviews and 25 participated in focus groups. Participants provided oral consent prior to participation and they received $25.00 for participation following completion of the focus group or interview. The rationale for using focus groups and individual interviews as the two qualitative data collection methods was that both methods allowed conducting an in-depth assessment of participants' comprehension of the toolkit's content.

The goals of the focus groups and individual interviews were: (1) to assess how much participants understood the information included in a consent and authorization form about a hypothetical study for patients with diabetes; (2) to obtain participants' input about how to improve comprehension; and (3) to explore the effectiveness of communication strategies other than reading in order to make the process of obtaining consent to participate in a research project meaningful to potential research participants. Although the consent form referred to a hypothetical study for patients with diabetes, the document did not have clinical content, and participant selection criteria did not include having diabetes. We informed participants that the consent form referred to a hypothetical study, and thus asked them to pretend that they had diabetes, and were being invited to participate in a study designed for people with diabetes.

Upon obtaining verbal instead of oral consent—in order not to exclude individuals with low literacy—to participate in the study, we gave participants a binder that contained the consent form and HIPAA authorization typed in non-glossy, high contrast paper with large print to avoid excluding individuals with low vision or other visual impairments. During both the focus groups and individual interviews we asked participants to review the text included in the

consent form and HIPAA authorization using a semi-structured interview format combined with cognitive interviewing (i.e., think-aloud) techniques. First, we asked participants to read the consent form and HIPAA authorization. For the focus groups, the facilitator asked the group to complete that task by taking turns reading the documents aloud.

Once participants read the documents, each statement included in them was reviewed following four steps. For the *first step*, the facilitator asked participants to review a statement on the document, and then asked, 'What does this statement mean to you?' Participants' perceptions of the phrase and what it meant to them were recorded, and follow-up questions were asked to further clarify participants' responses as necessary. If participants reported not knowing the meaning of a word, this was recorded. For the *second step*, the facilitator read the same statement aloud, and then asked questions about the statement (e.g., 'tell me in your own words what this means') followed by questions for clarification, if needed. After discussion of their perceptions, the *third step* involved the facilitator explaining what the statement meant and asking participants to think about how the wording could be improved to make the statement easier to understand or clearer (e.g., 'Which words need to be changed?' 'How can we say this better?'). Any different wording participants suggested in order to better or more easily explain the concept was noted. During *step four*, the facilitator presented textual variants of the same item to determine if other wording would be more appropriate.

Although the goal was not to reach consensus, if there was one relatively clear choice among focus group participants, that choice was marked as preferred. If the group agreed on multiple good options, they were noted and rank ordered per the group's preferences. This four-step procedure was repeated for each element in the sample informed consent and HIPAA authorization forms in order to address each of the elements of informed consent and authorization. At the end of the interviews and focus groups, participants were asked, ''tell me in your own words about the research project described in these documents.'' Based on feedback from participants, the text was adapted in an iterative fashion.

Results

After each participant of the first focus group took turns reading each statement included in the consent form and HIPAA authorization form and followed the four review steps described above, five of the seven members of this group could not explain the purpose of the documents in their own words. In contrast with participants in the first focus group, those in the other two focus groups as well as those who completed the individual interviews were able to formulate the documents' intent and content in their own words.

Most participants reacted *unfavorably* to the length of the documents. Even though the documents were written in plain language and followed low literacy principles, some participants deemed them too long. One participant provided the following recommendation: ''say what you need to say not using so many words.'' Another participant recommended not to use fine print since that kind of font ''is not trustworthy.'' All participants expressed their preference to have someone (e.g., researcher) going over the consent form with them. They also expressed the need to be given time and ''space'' to carefully review the consent form before making a decision about participating in the research study. This quote

illustrates what some participants deemed the appropriate way to handle the process of obtaining consent:

> "When someone hands you this document and sits in front of you while you are reading, it's intimidating and makes you nervous and you can't really read it well. So, we need time, and a properly designated space where we can read and sign it. If they can take it home and read it with a family member, it would be better. This way they'll feel more comfortable in discussing the document with THEIR family, rather than with a stranger. You're giving them options. You can say to them, "I can help you. You can do it by yourself, or you can take it with you." [It's important to give them options] because you are not aware of my skills yet."

Another participant expressed a similar opinion: "I would need to feel assured, just like we're doing here today. I would like to read it slowly, with various people, like we are doing here, and go step by step, in order for me to feel sure."

Most participants (in both focus groups and individual interviews) reported that they had never signed a consent or authorization form for research participation before. As they were going through the documents, it was evident that these individuals had difficulties understanding what it takes to participate in a research study and the purpose of going through the process of consenting. For example, participants had difficulties understanding the goals of the study as well as the research activities involved in the study described in the consent form. They also expressed limited understanding of the process of withdrawing from the research study and the use of personal health information obtained prior to withdrawing from the study.

Issues related to translation were also evident in the study. For example, participants provided feedback that some sentences contained words that were grammatically correct, but read like literal translations with a syntactic structure that followed English language rules, and thus had an impact on participants' comprehension. This finding underscores the importance of incorporating a pilot test to assess comprehension of the translated test prior to completing the final version of the translated during consent form.

Notable Areas of Complexity

Risk

A significant number of participants had difficulties understanding the purpose and intent of the section that described risks from participating in the research study. For example, many of them could not comprehend why the possibility that other people would learn about their participation in the study could be harmful. For other participants, the risk section explanation of mandatory reporting of crimes committed by participants instilled distrust among some participants, and made them consider not consenting to participate in the study described in the forms. In this regard, one participant said:

> "That is forewarning you that you're better off not opening your mouth.... And to tell lies, if you have any problem. Well, this part

> frightened me when I read it and it was the first thing I contested. The money part or the questions part was nothing. But this part tells me that I'm protected, but once I sign the document, I'm no longer protected. But, not only I am not protected for what I may say, but I can also end up in jail.''

Protected Health Information

After reading the documents some participants indicated that they would not provide consent to participate in the study described in the consent form and HIPAA authorization since they did not believe that their personal health information would be protected. The reason why they did not believe that their personal health information would be protected stemmed from what they described as confusing or contradictory information regarding the use this information. Participants who deemed this information confusing expressed a heightened level of distrust regarding the goals of the study and the need to sign a consent form. The following quotes illustrate the feelings participants expressed in this regard:

> ''I read in the document that my medical record is protected by a privacy law, but the minute I sign the consent, it is no longer protected. And it states that the information will be given to certain people. How am I going to be sure that people will not have access to that information? Besides, it states that I can stop the authorization at any moment, but that the information I have already provided will continue to be used. There really isn't any law that will protect me. In other words, what this is saying doesn't make sense.''

Another participant echoed the same feelings:

> ''Pardon me. It seems to me like this contradicts itself. It contradicts because first, it tells me that medical record information is protected under the Privacy Rule, and later on, it tells me that you are not responsible for protecting my information. It's not confusing. It's clear, but contradictory. We are no longer protected by the law because we are giving you authorization for other people to see it. They're going to be frightened. They [the research team] will use it for the study, for whatever they want. I repeat, it tells us, ''WE WILL DO OUR BEST . . .'' No one is giving me a written document or a guarantee. I'm not given a guarantee where it says, 'your information will be kept confidential with people of the study.'''

Other participants who felt that the documents felt short protecting their privacy, reported that since their participation in research could potentially produce knowledge that could benefit other people, they were willing to forgo full protection of their privacy in order to help others (see the Appendix for an example of toolkit's text). Finally, the study's results did not suggest that culturally-specific issues informed participants' willingness to participate in the hypothetical study described in the consent form.

Discussion

We undertook a series of cognitive interviews with iterative adaptation of consent and authorization documents to refine the template text for the AHRQ Informed Consent and HIPAA Authorization Toolkit. This activity was guided as well by a series of reviews conducted by a Delphi Panel of experts as well as by staff at the Office for Human Research Protections (OHRP) and the Office for Civil Rights (OCR). Findings from this study indicate that even when researchers follow recommendations designed to facilitate reading comprehension of consent forms—large font, plain language, shorter sentences, document written in respondent's preferred language, wide margins, and shorter paragraphs—comprehension is not guaranteed. While our experience provided evidence that improved readability is important, underlying conceptual complexity cannot be avoided in many instances. Some participants did not understand the content of the consent form even though they were able to read it. This appeared mostly for people with no prior experience with research and for those who have never seen a consent form. Indeed, lack of familiarity with the concepts appeared to be a more important barrier to comprehension than semantic or syntactic issues. Others understood most of the content, but were confused about the extent to which the consent form provided sufficient protections around the issues of confidentiality, risks and benefits. In addition to this, even though the documents had been edited to have fewer words than typically found in comparable documents, they were generally viewed as lengthy and wordy. It is clear that comprehension was not guaranteed through the application of technical document-focused strategies (e.g., large font, shorter sentences, etc.) designed to lower the document's reading level. This suggests that researchers should take additional steps to improve comprehension since it is not realistic to expect that reading a consent form will guarantee comprehension of the research project it describes (Sugarman & Paasche-Orlow, 2006).

Lack of familiarity with participation in a research study and lack of familiarity with providing informed consent and authorization appeared to play key roles in individuals' ability to understand what is being asked from them. Among those who understood both the process of participating in a research study and consenting to it via a written document, the language used in the consent form (even though it was considered written in ''plain language'') was deemed confusing and contradictory. Confusing and contradictory information in the HIPPA authorization form generated feelings of distrust among participants who demonstrated a high level comprehension of the document, and thus they felt that they did not have enough protections to be willing to sign the document.

The results of this study align with those of previous studies that have shown that comprehension of informed consent documents poses problems for many participants (Taub et al., 1986; Davis et al., 1998). Similar to the study described here, Davis and colleagues (2002) conducted a study to test whether the use of a simplified, illustrated consent document that followed low literacy recommendations would be easily understood by individuals with marginal-to-low reading skills. Although participants indicated that the simplified version was more conducive to participating in the study than a standard consent form, there was no difference in level of comprehension across the two consent form versions. They indicated that participants' reading levels appeared to be related to comprehension and thus questioned the use of written consent forms for low literacy individuals. Our results echo these earlier

findings, and demonstrate how basic research concepts can be challenging for people to comprehend even when they can read a document that describes those concepts. In addition to this, our study's findings regarding the absence of culturally informed bias towards consenting to participate in a research project compare with findings from another study conducted with Mexican Americans. That study focused on their recognition of the risks and ethical issues associated with enrollment in genetic family studies and explored how this recognition affects their informed and voluntary participation. The study found that participants did not recognize and tended to underestimate the social and cultural risks associated with their participation in genetic family studies (Arar et al., 2005).

The study provided the opportunity to further refine the Toolkit by improving the Spanish and shorten the text. Findings from this study also provide further evidence that energy needs to be put into the process of obtaining consent and authorization from potential research participants, since simply translating documents from one language into another is not enough to achieve comprehension. Institutional Review Boards, in their effort to protect human participants, need to provide specific guidelines to researchers who recruit individuals who speak languages other than English, have low literacy skills, and are unfamiliar with both research participation and informed consents. For those who speak languages other than English, Institutional Review Boards' requirement of conducting back-translation procedures to verify translation addresses the issue of content accuracy but not the issue of content comprehension. Back-translation is considered a rough and mechanistic assessment of translated text (Harkness et al., 2004), and thus falls short determining the translation's quality in terms of the level of comprehension on the part of the audience who will be using the translated text. In addition to that, a back-translation of a literal translation may simply replicate flaws present in the original text. In this sense, a back-translation could fall short as a tool to assess the adequacy of the translated text (Blais and Gidengil 1993). IRBs should provide templates of consent forms in languages other than English that are not only translated by professional translators but pilot tested for comprehension with individuals who speak those languages and who have different levels of literacy and familiarity with research and informed consent. Implementing these steps could lead to language-specific consent forms that successfully capture and clearly communicate the intended concepts to individuals with little or no familiarity with participation in research.

Since the data obtained from this research was communicated to AHRQ prior to the publication of this manuscript, the Toolkit—readily available to the research community—recommends that investigators obtaining consent from potential research participants implement interactive strategies to confirm participant's comprehension of the consent form. Our research design allowed us to do this, and it provided an opportunity to learn how participants interpreted the documents' content (i.e., confidentiality, risk) and further details of their decision-making process (e.g., forgoing privacy in order to help others through their participation in research). We recognize that, realistically, most researchers cannot spend unlimited amount of time obtaining consent from potential participants. However, in order to make comprehension of both the consent form and the research study in question a universal guideline or standard, Institutional Review Boards should also require that researchers complete training on interactive consent processes designed to address the informational needs of individuals with low literacy levels and limited or no

experience with research participation (Sudore, 2006). The training should include the use of prescreening protocols aimed at explaining the purpose of the research study to potential participants prior to obtaining consent as well as the use of probes and teach-back methods as tools to inform about consent and authorization and to verify comprehension. Such procedures are needed not only to conform to regulatory guidelines and ethical norms, but also, to improve access for a broad spectrum of potential participants to research protocols that they might otherwise avoid simply due to misunderstanding. The failure to enroll participants with low health literacy can threaten the validity of many research protocols. Finally, our findings suggest the need for IRBs and researchers alike to consider adding to consent forms a brief introductory statement explaining what participation in research is and why potential participants need to be informed about what is involved in the research study before they consent to join the study.

References

Arar, N. H., Hazuda, H., Steinbach, R., Arr, M. Y., & Abboud, H. E. (2005). Ethical issues associated with conducting genetic family studies of complex disease. *Annals of Epidemiology*, *15*(9), 712–719.

Blais, A., & Gidengil, E. (1993). Things are not always what they seem: French-English differences and the problem of measurement equivalence. *Canadian Journal of Political Science*, *26*, 541–555.

Coyne, C. A., Xu, R., Raich, P., Plomer, K., Dignan, M., Wenzel, L. B., et al. (2003). Randomized, controlled trial of an easy-to-read informed consent statement for clinical trial participation: A study of the Eastern Cooperative Oncology Group. *Journal of Clinical Oncology*, *21*(5), 836–842.

Davis, T. C., Holcombe, R. F., Berkel, H. J., Pramanik, S., & Divers, S. G. (1998). Informed consent for clinical trials: A comparative study of standard versus simplified forms. *Journal of National Cancer Institute*, *90*, 668–674.

Davis, T. C., Williams, M. V., Marin, E., Parker, R. M., & Glass, J. (2002). Health literacy and cancer communication. *CA Cancer Journal for Clinicians*, *52*, 134–149.

Flory, J., & Emanuel, E. (2004). Interventions to improve research participants' understanding in informed consent for research: A systematic review. *Journal of American Medical Association*, *292*(13), 1593–1601.

Harkness, J. B., Pennell, E., & Schoua-Glusberg, A. (2004). Survey questionnaire translation and assessment. In S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin, & E. Singer (Eds.), *Methods for Testingand Evaluating Survey Questionnaires* (pp. 453–473). Hoboken, NJ: John Wiley and Sons.

Jenkins, V. A., Anderson, J. L., & Fallowfield, L. J. (2010). Communication and informed consent in phase 1 trials: A review of the literature from January 2005 to July 2009. *Support Care Cancer*, Mar 4, 1115–1121.

Kang, E. Y., Fields, H. W., Kiyak, A., Beck, F. M., & Firestone, A. R. (2009). Informed consent recall and comprehension in orthodontics: Traditional versus improved readability and processability methods. *American Journal Orthodontic and Dentofacial Orthopedics*, *136*(4), 488.e1–13; discussion 488–489.

Paasche-Orlow, M. K., Taylor, H. A., & Brancati, F. L. (2003). Readability standards for informed-consent forms as compared with actual readability. *New England Journal of Medicine*, *348*, 721–726.

Paris, A., Brandt, C., Cornu, C., Maison, P., Thalamas, C., & Cracowski, J. L. (2010). Informed consent document improvement does not increase patients' comprehension in biomedical research. *British Journal of Clinical Pharmacology*, *69*(3), 231–237.

Raich, P. C., Plomer, K. D., & Coyne, C. A. (2001). Literacy, comprehension, and informed consent in clinical research. *Cancer Investigation*, *19*(4), 437–445.

Sudore, R. L., Landefeld, C. S., Williams, B. A., Barnes, D. E., Lindquist, K., & Schillinger, D. (2006). Use of a modified informed consent process among vulnerable patients: A descriptive study. *Journal of General Internal Medicine*, *21*, 867–873.

Sugarman, J., McCrory, D. C., Powel, D., Krasny, A., Adams, B., Ball, E., & Cassell, C. (1999). Empirical research on informed consent. An annotated bibliography. *Hastings Center Report*, *29*(1), S1–S42.

Sugarman, J., & Paasche-Orlow, M. (2006). Confirming comprehension of informed consent as a protection of human subjects. *Journal of General Internal Medicine*, *21*(8), 898–899.

Taub, H. A., Baker, M. T., & Sturr, J. F. (1986). Informed consent for research: Effects of readability, participant age, and education. *Journal of the American Geriatric Society*, *34*, 601–606.

Appendix: Example of AHRQ's Informed Consent and HIPAA Authorization Toolkit

English version	Spanish version
We are asking you to be in a research study.	Le estamos pidiendo que participe en un estudio.
You do not have to be in the study.	Usted no tiene que participar en el estudio.
If you say yes, you can quit the study at any time.	Si dice que sí, puede dejar de participar en el estudio en cualquier momento.
Please take as much time as you need to make your choice.	Por favor tome todo el tiempo que necesite para decidir.
Your medical care will not change in any way if you say no.	Su atención médica no cambiará de manera alguna si dice que no.
Why are you doing this research study?	**¿Por qué se está haciendo este estudio de investigación?**
We want to learn more about how to help people who have [insert condition]. This study will help us learn more about [insert specifics]. We are asking people like you who have [insert condition] to help us.	Queremos saber más sobre cómo ayudar a las personas que tienen [inserte condición]. Este estudio nos ayudará a aprender más sobre [provea información específica]. Les estamos pidiendo a personas como usted que tienen [inserte condición] que nos ayuden.
Is there any way being in this study could be bad for me?	**Participar en este estudio, ¿pudiera ser malo para mí, de alguna manera?**
Yes. There is a chance that:	Sí. Hay una posibilidad de que:
• The questions could make you sad or upset. • Someone could find out that you were in the study and learn	• Las preguntas le hagan sentirse triste o sentirse mal. • Alguien pudiera enterarse de que usted participó en este estudio y

(*Continued*)

Appendix: Continued

English version	Spanish version
something about you that you did not want others to know. • You could have a legal problem if you told us about a crime such as child abuse [list other mandatory reporting required in your state] that we have to report.	llegar a saber algo sobre usted que usted no quería que supiera. • Usted podría tener un problema legal si nos cuenta sobre un delito, como el abuso de niños [proporcione una lista de asuntos de notificación forzosa que se exijan en su estado], que tenemos que reportar.
We will do our best to protect your privacy.	Haremos todo lo posible para proteger su privacidad.
[**Note to researcher:** Insert details on additional risks if relevant to the study, such as: You could have a legal problem if someone outside the study found out that you did something illegal.] [Provide details regarding accommodation or referrals (e.g., for counseling) if relevant to the study.]	[**Nota para el investigador:** Provea detalles sobre riesgos adicionales si son relevantes para el estudio, tales como un problema legal si alguien fuera de este estudio se enterara que usted hizo algo ilegal.] [**Nota para el investigador:** Provea detalles sobre asistencia o referidos (por ejemplo, consejería) si es relevante para el estudio]

beneficial in terms of CV outcomes, but this was presumably confounded by the disadvantages of the atenolol-based regimen.

However, we did not focus on the variable predictability of HR because that was not the purpose of the study. We wanted to know whether having a higher baseline HR attenuated the superior effects on major CV events of the amlodipine-based compared with atenolol-based regimen. We could find no evidence of any such attenuation, and hence we believe that an increased baseline HR should not be an indication for preferential use of beta-blockade in hypertensive populations without coronary heart disease. Even if baseline HR had not predicted CV outcomes in the ASCOT study, we believe that the same conclusion should be drawn.

***Neil R. Poulter, MB, MSc**
Joanna E. Dobson, MSc
Peter S. Sever, PhD
Björn Dahlöf, MD, PhD
Norm R. C. Campbell, MD

*International Centre for Circulatory Health
Imperial College London
59 North Wharf Road
London, W2 1PG
United Kingdom
E-mail: n.poulter@imperial.ac.uk

doi:10.1016/j.jacc.2009.11.038

REFERENCES

1. Poulter NR, Dobson JE, Sever PS, et al. Baseline heart rate, antihypertensive treatment, and prevention of cardiovascular outcomes in ASCOT (Anglo-Scandinavian Cardiac Outcomes Trial). J Am Coll Cardiol 2009;54:1154–61.
2. Hansen TW, Thijs L, Boggia J, et al. Prognostic value of ambulatory heart rate revisited in 6928 subjects from 6 populations. Hypertension 2008;52:229–35.
3. Bangalore S, Sawhney S, Messerli FH, et al. Relation of beta-blocker-induced heart rate lowering and cardioprotection in hypertension. J Am Coll Cardiol 2008;52:1482–9.
4. Kolloch R, Legler UF, Champion A, et al. Impact of resting heart rate on outcomes in hypertensive patients with coronary artery disease: findings from the INternational VErapamil-SR/trandolapril STudy (INVEST). Eur Heart J 2008;29:1327–34.
5. Poulter NR, Wedel H, Dahlöf B, et al. Role of blood pressure and other variables in the differential cardiovascular event rates noted in the Anglo-Scandinavian Cardiac Outcomes Trial-Blood Pressure Lowering Arm (ASCOT-BPLA). Lancet 2005;366:907–13.

Bleeding Risk on Warfarin Among Elderly Patients With Atrial Fibrillation

Poli et al. (1) observed different rates of major hemorrhage between patients younger than 80 years of age and 80 years of age and older. As discussed by the authors, these rates differed considerably from the rates of major hemorrhage observed among similarly aged cohorts by Hylek et al. (2). We want to highlight an important methodological issue in the authors' calculation of event rates. The authors state, "the overall exposure to warfarin for each patient was calculated in relation to aging, before and after his/her 80th birthday." Thus, the authors allowed crossover of prevalent warfarin survivors from the younger cohort to the age ≥80 years inception cohort. At the time of enrollment, the baseline age ≥80 years cohort numbered 180 patients. Yet, in Table 1 of their article (1), the authors provide baseline characteristics for 327 patients in the age ≥80 years group. The 2 age inception cohorts are distinct and should contribute person-years exclusively to their baseline assignment. Given this methodological error, the rates of hemorrhage provided for the 2 inception cohorts are flawed. The reader is also unable to compare baseline characteristics between the younger and age ≥80 years inception cohorts because the authors permitted crossover of 157 patients. In addition, the observation period in the study by Hylek et al. (2) was intentionally truncated at 1 year to provide the first-year experience on warfarin. Calculation of adverse event rates over years tends to enrich the person-year denominator with "survivors" because bleeding rates are highest in the first 90 days of warfarin therapy. To accurately report rates of major hemorrhage among elderly individuals newly starting warfarin, the authors need to recalculate the bleeding rates according to baseline group assignment without crossover between the groups. To enable comparison of the 2 studies, events and person-years of observation would need to be restricted to the patients' first year of therapy. The anticipated results would be higher bleeding rates and deterioration of time "in-range" as reported.

***Andrew Cowan, MD**
Elaine M. Hylek, MD, MPH

*Department of Medicine
Boston University Medical Center
72 East Concord Street
Evans 124
Boston, Massachusetts 02118
E-mail: andrew.cowan@bmc.org

doi:10.1016/j.jacc.2009.11.037

REFERENCES

1. Poli D, Antonucci E, Grifoni E, Abbate R, Gensini GF, Prisco D. Bleeding risk during oral anticoagulation in atrial fibrillation patients older than 80 years. J Am Coll Cardiol 2009;54:999–1002.
2. 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;115:2689–96.

Reply

We studied a cohort of atrial fibrillation patients on oral anticoagulant treatment for stroke prophylaxis (1). Our patients were routinely followed by the Anticoagulation Clinic of our institution with a median follow-up of 2.7 years, and some of them for as long as 13 years. At the beginning of warfarin treatment, the mean age of our cohort was 75 years; therefore, many patients reached the age of 80 years during follow-up. As stated in the article, we decided to analyze the occurrence of adverse events in relation to aging to evaluate whether aging itself could be correlated with an increase in bleeding risk that exceeds the advantages of stroke prevention. In reporting clinical characteristics of patients, we

The Laryngoscope

Contemporary Review

Head and Neck Sequelae of Torture

Sondra S. Crosby, MD; Shaulnie Mohan, MD; Christina DiLoreto, MD; Jeffrey H. Spiegel, MD, FACS

Objectives/Hypothesis: To increase awareness of torture among otolaryngologists, and to describe methods and complications of head and neck torture.

Study Design: Retrospective review.

Methods: Five cases of survivors of torture were evaluated in an otolaryngology practice in an urban hospital setting.

Results: The subjects presented with widely variable symptoms and physical manifestations related to the head and neck as a result of torture, in addition to psychiatric disease. Documentation of head and neck findings were essential to the asylum claim.

Conclusions: Otolaryngologists serving immigrant and refugee populations must be familiar with methods and manifestations of torture involving the head and neck.

Key Words: Head and neck, torture, methods of torture, trauma.

***Laryngoscope,* 120:414–419, 2010**

INTRODUCTION

The UN Convention Against Torture defines torture as: "any act by which severe pain or suffering, whether physical or mental is intentionally inflicted on a person for such purposes as obtaining from him or her or a third person information or a confession, punishing him for an act he or a third person has committed, or intimidating or coercing him or a third person, or for any reason based on discrimination of any kind, when such suffering is inflicted by or at the instigation of or with the consent or acquiescence of a public official or other person acting in an official capacity. It does not include pain or suffering arising from, inherent in or incidental to lawful sanctions."[1]

From the Department of Medicine (S.S.C.), Boston University School of Medicine; the Department of Health Law, Bioethics and Human Rights, Boston University School of Public Health (S.S.C.), Boston University School of Medicine (S.M., C.D.), and the Department of Otolaryngology–Head and Neck Surgery, Boston University School of Medicine (J.H.S.), Boston, Massachusetts, U.S.A.

Editor's Note: This Manuscript was accepted for publication August 19, 2009.

Send correspondence to Sondra S. Crosby, MD, Section of General Internal Medicine, Boston Medical Center, Crosstown 2, 801 Massachusetts Ave., Boston, MA 02118. E-mail: scrosby@bu.edu

DOI: 10.1002/lary.20731

In 2008, Amnesty International Report reported that torture and ill treatment occurred in 81 countries[2]; however, torture is commonly under-reported. At the end of 2007, there were 31.7 million people of concern to the UNHCR (refugees, asylum seekers, displaced persons, and others), 365,103 of whom reside in the United States.[3] The United States was the main receiving country for asylum applications, with 50,000 lodged in 2007.[3] A refugee is defined as an alien unwilling to return to his or her country of origin "because of persecution or a well-founded fear of persecution on account of race, religion, nationality, membership in a particular social group, or political opinion."[4] Asylum seekers meet the same criteria as refugees; however, they apply for asylum in the host country. This distinction is important because asylum seekers may be undocumented, and are an extremely vulnerable population. The prevalence of torture among refugee groups resettled in the United States varies with the population studied. For example, Jaranson et al. reported that the prevalence of torture in Somali and Oromo refugees in Minneapolis-St. Paul was 36% and 55%, respectively.[5] In a clinic-based survey of foreign-born patients presenting to an urban primary care clinic in Boston, 11% reported a history of torture.[6] Associated with a higher risk of torture were recent arrivals to the United States from the African and Asian continents.

When survivors of torture do seek medical care, it may be their first interaction with the healthcare system. Survivors may choose not to disclose their experiences due to fear of putting themselves and families in further danger, impairment of memory resulting from torture, cultural sanctions, or simply as a coping strategy.[7] Survivors of torture may have difficulty developing trust with healthcare providers, as one third to one half of survivors report that physicians oversaw their torture.[8] In one report, 6.6% (8/121) of foreign-born patients polled in a large urban clinic in the United States reported a history of torture, and none of these patients were recognized as torture survivors by the treating physicians.[9]

Torture is an assault of a person's mind, body, and sense of security, and may cause long-term physical and psychological effects. It is one of the most traumatizing of human experiences. Torture methods are often devised so that they leave minimal or no physical signs of torture after the fact, thus it is important for

providers to consider the possibility of torture in the appropriate clinical setting.[10]

Mechanisms of torture specific to the head and neck include, but are not limited to, beatings resulting in lacerations and facial fractures; cutting; burning (thermal and chemical); application of electric shocks to the tongue, lips, and ears; dental trauma; tympanic membrane rupture from deliberate hard slaps to the ears (known as teléfono); barotrauma from loud noises resulting in hearing loss; and near asphyxiation (e.g., waterboarding).[10] The physical symptoms and complications resulting from these forms of torture may require specialized evaluation and treatment by an otolaryngologist.

The objective of this report is to increase awareness of torture among otolaryngologists. This paper will specifically review prevalence of and methods of torture involving the head and neck. It will emphasize the importance of establishing rapport, obtaining a social history, and determining the method of torture to appropriately diagnose and adequately treat survivors of head and neck torture.

MATERIALS AND METHODS

A series of case reports of asylum seekers referred to the Boston Medical Center Department of Otolaryngology from the Boston Center for Refugee Health and Human Rights (BCRHHR) between May 2004 and December 2008 for evaluation and treatment of injuries to the head and neck as a result of torture in their home countries is presented. Cases included women and men $\geq$18 years of age who were survivors of torture by the United Nations definition. We describe the methods of torture, symptoms, physical exam, diagnosis, and treatment of these cases. Methods and manifestations of head and neck torture are reviewed. The institutional review board at Boston Medical Center approved the study.

RESULTS

Case 1

Patient 1 is a male in his mid-30s who fled from a Middle Eastern country after being persecuted for his religious affiliation. He was kidnapped and beaten for several hours, including sustaining blunt trauma to the head resulting in loss of consciousness. He was referred to the BCRHHR by his attorney for a forensic evaluation as part of his application for political asylum. Psychological manifestations of his torture included post-traumatic stress disorder (PTSD) and major depressive disorder. The patient complained of episodic dizziness lasting days to weeks, causing difficulty with ambulation and associated otalgia necessitating multiple visits to the emergency department. He did not relate the history of beating to his current symptoms, and was initially fearful of disclosing his symptoms. He worried that the gait imbalance that he experienced would cause people to suspect he was drinking alcohol, which was culturally taboo. The patient was referred by his primary care doctor to the otolaryngology clinic, where on physical examination, the patient was found to have normal cerebellar function, normal Romberg standard, and an abnormal Romberg tandem. With evaluation of gait, the patient was found to have normal gait with open eyes, but a slower, broad-based gait with closed eyes. Dix-Hallpike exam revealed right side down horizontal nystagmus for 10 seconds without latency, followed by rebound nystagmus on recovery of position. Otoscopic exam revealed normal external auditory canals and tympanic membranes bilaterally, however fistula test on the left was equivocal. Post-traumatic disequilibrium secondary to left oval window fistula was suspected, and vestibular and audiometric testing was ordered. Findings from the additional testing included left sensorineural hearing loss with otherwise negative electrocochleography (audiometric and vestibular). Because an oval window fistula was still suspected, the patient was taken to the operating room for exploratory tympanotomy, which revealed pooling of fluid around the oval window, confirming the diagnosis of an oval window fistula. Fat harvested from the pretragal tissue was used to repair the defect, and the graft was secured with Gelfoam. The patient is seen regularly for follow-up, and he has noted no return of the vertiginous symptoms since the repair. An affidavit in support of his asylum claim documenting his otolaryngolic injury was written by his primary care physician with input from the otolaryngologist, and the patient was granted asylum.

Case 2

Patient 2 is a female in her late 50s who fled from Central Africa after being beaten unconscious on two separate occasions for being a member of an opposition political party. Her primary complaint on presentation was excessive tearing, rhinorrhea, and a collapsed left nostril. She could not recall with what she was beaten, but she had significant scarring over her face and legs. She was referred to otolaryngology for further evaluation and treatment by her primary care physician at the BCRHHR. On physical exam, she had scarring of the left eyelid with retraction and mechanical lagophthalmos, multiple scars on her face with right medial canthus splitting, palpable old fractures at the left orbital rim, abnormal occlusion, thinned tissue over the right frontal bones, and decreased mobility of the right forehead. She was also noted to have nasal dorsum flattening, near total stenosis of the left nasal vestibule, and nasolabial fold scarring with displacement of the nasal spine. Neurological exam showed decreased mobility of the left frontalis muscle and paresthesias in the left ophthalmic and maxillary distributions. The patient was offered surgical repair of the nasal vestibular stenosis in addition to scar revision of the lower eyelid and right alar area. After discussing her options, the patient chose to defer treatment until the resolution of her legal issues. She has yet to return to clinic for follow-up. She suffers from cognitive impairment as a result of the head injury, in addition to PTSD and major depressive disorder, all of which complicate her medical care and her legal case. An affidavit documenting her head and neck injuries was presented to the court, and her case for political asylum was granted.

Case 3

Patient 3 is a male in his late 40s from Eastern Africa who was detained and tortured in his home country on multiple occasions because of his ethnicity. His people were persecuted, their villages pillaged, and the children abducted to serve in the army. He was arrested by the government and tortured (including beatings, electric shocks, and genital trauma). He sustained multiple injuries, including avulsion of his auricle with a sharp object and being forced to swallow it. He was seen by a primary care doctor in the BCRHHR, and referred to the otolaryngology clinic. He initially reported that his injuries resulted from a motor vehicle accident, but later admitted that his auricular injury actually occurred while confined in a torture camp. He was reluctant to disclose his torture history because of shame. His physical examination revealed a severed left ear with approximately two thirds of the superior helix and triangular fossa absent. The superior aspect of his external auditory canal was stenotic. The soft tissue was intact with a straight line laceration obliquely between the residual cartilage lobule and posterior auricular skin. The patient was felt to be a good candidate for reconstructive surgery. He underwent auricular reconstruction with a Medpor framework, temporalis fascia flaps, and a split thickness skin graft from the lower extremity. This patient suffers from PTSD and generalized anxiety disorder, requiring specialized psychiatric care. An affidavit documenting his head and neck injuries was submitted to the court, and he was granted political asylum.

Case 4

Patient 4 is a female in her mid-20s from a Middle Eastern country, who was strangled to unconsciousness and left in a bathtub where hot water scalded and burned her face and chest in a gender-based attack. It is not known whether chemicals were also used to enhance the extensive burn injury. When she awoke, she reported "there was no skin left on my face, it came off like soft cheese." She was hospitalized and did not regain consciousness for several months. She has undergone multiple reconstructive surgeries in several countries, and is seeking political asylum in the United States. Her examination revealed extensive scarring and contractures on her face, chest, and neck including extensive scarring of her eyelids. Prior surgeries including significant split thickness skin grafting to restore epithelial coverage to areas of debrided skin.

Case 5

Patient 5 is a male in his early 20s from Western Africa who was assaulted by a local government supported vigilante group that was heavily involved with theft, arson, and property damage. When he refused to join the organization, they retaliated by pouring acid on his face while he was sleeping.

He was seen by a primary care doctor in the BCRHHR and referred to the otolaryngology clinic for further examination of his injuries. He initially reported that his injuries had occurred as a result of a property dispute and that his brother had been responsible for the attack. Later, it was revealed his injuries were the result of vigilante violence. At the time the patient presented to the clinic, he had undergone partial auricular reconstruction at another institution. A Medpor graft had been placed to expand the right ear with a temporoparietal fascial flap 2 years earlier. He presented to us desiring to complete reconstruction. On physical exam, extensive burns were noted involving the right side of his face extending posteriorly to involve his scalp and inferiorly to involve the superior aspect of his lips. He had significant scarring and retraction of the upper and lower lids of his right eye, with opacification and injury to his cornea. He admitted to only being able to see shadows on the right side. There was a scar noted across the front of his scalp from a tissue expander. Examination of his right ear revealed cartilaginous structure under the scar and skin. No posterior sulcus or external auditory canal was visualized. After preliminary imaging was done, the patient was found to be a good candidate for completion of the reconstructive surgery. The patient underwent stage 3 reconstruction of his right ear with an otoplasty, lateralization of his auricle, and split thickness skin graft from his upper extremity. Postoperative follow-up revealed a patent right external auditory canal with an intact and mobile tympanic membrane. The patient was pleased with the outcome and reported improved hearing on his right side. His asylum claim is pending.

DISCUSSION

Head and neck trauma is frequently reported by torture survivors. Beating to the head is a common form of torture, with one report citing 73% of 200 survivors reported severe beatings to the head. In this series, vertigo was reported in 20%, of whom 87% had been beaten on the head.[11] Another study of 63 consecutive torture survivors in Denmark reported that 95% of the subjects had been beaten on the head and neck, and that 17% reported being subjected to teléfono. The most common presenting symptoms in this cohort were tinnitus, decreased hearing, impaired air passage through the nose, and dizziness. There was a significant relationship between teléfono torture and tinnitus.[12] Other common techniques of torture are listed in Table I.[13,14] The prevalence of torture survivors at our institution presenting with head and neck injuries requiring evaluation by an otolaryngologist is not known. The five cases in this report presented to care between May 2004 and December 2008; however, this sample does not include all survivors of torture with head and neck torture. Of 20 forensic evaluations completed by the first author (S.C.) in 2008, 12 (60%) had documented injuries involving the head and neck. These injuries included blunt trauma resulting in facial lacerations, loss of consciousness, a mandible injury requiring surgical repair, vertigo, dental avulsions; cutting wounds with a blade or knife to the face or lips; administration of electric shocks to the lips; extensive burns to the head and neck; and strangulation

TABLE I.
Most Common Types of Physical Torture[13]

The following is a list of the most common types of tortures experienced by survivors of torture:
Blunt trauma
Penetrating injuries
Suspension
Burns: chemical and thermal, cold and heat
Asphyxiation: wet, dry, chemical
Electric shocks
Forced human experimentation
Traumatic removal of tissue and appendages
Extreme physical conditions
Sexual trauma

injuries. Table II lists additional injuries and complications that can occur as a result of torture involving the head and neck.

Head and neck examination of the torture survivor should focus on the following findings:

Face

Fractures or soft tissue injuries can occur as a result of beatings to the head and neck. Scarring from lacerations, cutting wounds, or burns may be present after healing of the acute injury. Scars from trace electrical burns may be present on the ears, mouth, lips, or tongue. In addition, scars may be present on the lips or tongue from involuntary biting during the administration of electric shocks.

Nose/Paranasal Sinus

Examination of nasal structures and paranasal sinuses may reveal obvious pathology, such as a nasal fracture or a more subtle complication of torture, such as chronic sinusitis. Chronic sinusitis has been hypothesized to result following certain methods of torture. For example, Bangladeshi survivors who had been subjected to water treatments commonly reported complaints of chronic sinusitus.[15] Water treatment involves deluging a person's nasal cavity with hot or polluted water.[10,15] A torture technique known as submarino, which can be similar to water treatment, is a more common form of torture utilizing water. Wet submarino involves immersion of the head into fluids contaminated with urine and feces and drowning or near drowning a person. Aspiration of contaminated fluids leading to pneumonia is an immediate concern. With dry submarino, respiration is prevented by placing a plastic bag over the head, forcefully closing the oral and nasal passages, applying pressure over laryngeal structures, or forcing aspiration of various dusts, chemicals, or peppers.[10] Survivors may present with various complications ranging from facial edema/congestion, petechiae, bleeding, pneumonia, sinusitis, or other infection of the head and neck.[14]

Ears

Scarring on the tympanic membrane or any deformities to the external canal may be a significant finding related to a patient's history of torture. Teléfono, exercised in many regions, is a hard slap of the palm to the ears causing a rapid increase in pressure in the ear canal resulting in perforation of the tympanic membrane.[10] As small perforations heal quickly, these injuries may go unrecognized. Survivors may have been exposed to continuous loud music or noise at close range with resultant hearing loss. Survivors may also present with fluid in the middle ear, hearing loss, ossicular chain disruption, or fractures.

Oral Cavity, Pharynx, and Larynx

Dental torture, such as manual extraction of teeth and application of electrical current to the teeth, may cause a variety of complaints, including tooth avulsion or fracture. Survivors may present with mucosal swelling, bleeding, muscles spasms, and trismus or other limitations of jaw movement. Temporomandibular joint syndrome may develop as a result of beatings to the jaw.[10] Laryngeal damage may be the consequence of strangulation. Asphyxiation as a result of this pressure is another common method of torture that has a variable presentation, and may leave no physical sequelae. Laryngeal injuries may progress days later, leading to

TABLE II.
Head and Neck Torture.

Blunt trauma
Loss of consciousness
Facial lacerations, scarring
Facial fractures, deformities
Nasal fracture, deviated septum—Stenosis nasal vestibule and impaired air passage
Post-traumatic disequilibrium secondary to oval window fistula
Tympanic membrane perforation (teléfono)
Tinnitus (teléfono)
Temporomandibular joint syndrome
Parotid duct injuries
Sharp instrument wounds
Avulsion of auricle, lips
Cutting wounds to face, lips
Burns
Acid or hot water burns to face
Strangulation injuries
Laryngeal damage, vocal cord dysfunction
Electric shocks
Simulated drowning or asphyxiation (wet or dry submarino)
Sinusitis/infections, facial edema/congestion
Otitis externa
Hearing loss
Barotrauma from exposure to loud noises
Dental
Manual extraction of teeth

TABLE III.
Questions When Inquiring About Experiences of Torture.

In what country were you born?
Can you tell me what made you leave your country?
Have you ever had problems because of your culture or tribe? Your political belief? Your religion? Your gender?
Have you ever been arrested or put in jail?
Have you ever been beaten or attacked by soldiers, police, or rebel groups?
Have you ever seen or heard others being beaten or attacked?
Have any members of your family been arrested or attacked because of their culture, tribe, political beliefs, or religion?
If a history of torture or trauma is elicited, what problems are you having now from being beaten or attacked?

airway obstruction or vocal cord dysfunction.[16] If cerebral anoxia is severe, the survivor may suffer loss of consciousness, seizures, and incontinence.[17] In addition, long-lasting cognitive dysfunction may result from anoxic brain injury.

Prior otolaryngolic injuries may go unrecognized if there are no obvious physical findings. Specific techniques are often utilized to limit long-term evidence of injury and thus, a detailed history of the injury, symptoms, and healing process becomes crucial for documentation.[10] Furthermore, survivors may be tortured early in their imprisonment, and by the time they are released from detainment and/or seek medical care in the country of asylum, physical signs of torture may have healed. Survivors may not be forthcoming about their history, and although many patients presenting to clinic will have visible or easily recognizable injuries, others will present with subtle complaints such as vertigo and sinus pain. Radiologic studies (plain films, computed tomography, magnetic resonance imaging, and scintigraphy) may be useful in the evaluation of torture survivors who have sustained head and neck torture, and results can be correlated with history and physical examination findings.[18] Scintigraphy can sometimes support a history of beatings even years after they occur.[19] Evaluation of survivors of torture will require both awareness and clinical judgment.

Several risk factors have been identified for torture including: refugee or asylee status, immigrant from a country where there has been civil war or a totalitarian or military regime, history of arrest or detention or prisoner of war, relative of a torture survivor, or leader of an opposition organization.[20] When treating members of these groups, otolaryngologists should consider that presenting symptoms may be the result of torture. This knowledge can inform the sensitivity with which the history and exam is conducted. As a result of the circumstances of the injuries, it may be necessary to proceed more slowly, and take breaks through the interview or examination. Table III lists questions that can be helpful when inquiring about experiences of torture.[21] In addition to the physical manifestations of torture, the prevalence of PTSD and depression are high among torture survivors.[22] These symptoms may include insomnia, nightmares, flashbacks, intrusive thoughts, and avoidance of triggers. Talking about one's history can potentially elicit an emotional response. Being vigilant in observing for signs of hyperarousal or re-experiencing (flashbacks, dissociation) is necessary so as to adjust the flow of the examination. Awareness of mental health issues, specifically PTSD, is an important preoperative consideration. One case report describes a torture survivor who had a postoperative course complicated by an emergence flashback. The patient was delirious, recapitulating her previous torture experience, and required relocation to a private room with a female nurse and implementation of grounding techniques.[23] Creation of a safe environment is an important consideration when caring for survivors of torture in the medical setting.

Providers should be aware that centers specializing in the care of survivors of torture exist and are available for referral, in addition to mental health providers who specialize in the care of trauma survivors. The BCRHHR works with survivors of torture and refugee trauma. The center is a member of the National Consortium of Torture Treatment Programs,[24] and operates as an interdisciplinary collaboration providing comprehensive medical, mental health, and dental care coordinated with legal and social services to individuals. A web course on caring for survivors of torture is available at www.bcrhhr.org.[25]

CONCLUSION

There are few reports in the literature that specifically address the topic of torture involving the head and neck. The manifestations are broad and depend on the torture method utilized. Otolaryngologists who see immigrant and refugee populations will see patients with sequelae of head and neck torture. Physicians may lack awareness of, or feel discomfort asking about torture. It is important that otolaryngologists are trained to recognize the signs and symptoms of torture, and make appropriate referrals to mental health specialists or specialized torture treatment programs. In addition, otolaryngologists may be asked to document injuries as part of the patient's application for asylum.

Acknowledgments

The authors would like to thank Dr. Lin Piwowarczyk of the Boston Center for Refugee Health and Human Rights for providing her exceptional expertise and commentary to the manuscript.

BIBLIOGRAPHY

1. Office of the High Commission for Human Rights. Convention against torture and other cruel, inhuman or

degrading treatment or punishment. United Nations Web site. Available at: http://www2.ohchr.org/english/law/cat.htm. Accessed June 2, 2008.
2. Amnesty International Report 2008. Report 08: at a glance. Amnesty International Web site. Available at: http://archive.amnesty.org/air2008/eng/report-08-at-a-glance.html. Accessed July 29, 2009.
3. Diallo K, Chabake T.2007 UNHCR statistical yearbook: trends in displacement, protection, and solutions. UNHCR Web site. Available at: http://www.unhcr.org/statistics. Accessed July 29, 2009.
4. The Refugee Act of 1980, Title IV, ch 2, §411–414.
5. Jaranson JM, Butcher J, Halcon L, et al. Somali and Oromo refugees: correlates of torture and trauma history. *Am J Public Health* 2004;94:591–598.
6. Crosby SS, Norredam M, Paasche-Orlow MK, et al. Prevalence of torture survivors among foreign-born patients presenting to an urban ambulatory care practice. *J Gen Intern Med* 2006;21:764–768.
7. Mollica RF, Caspi-Yavin Y. Measuring torture and torture-related symptoms. *Psychol Assess* 1991;3:581–587.
8. Miles SH, Freedman AM. Medical ethics and torture: revising the Declaration of Tokyo. *Lancet* 2009;373:344–348.
9. Eisenman DP, Keller AS, Kim G. Survivors of torture in a general medical setting: how often have patients been tortured, and how often is it missed? *West J Med*2000; 172:301–304.
10. Iacopino V, Allden K, Keller A, eds. *Examining Asylum Seekers: A Health Professional's Guide to Medical and Psychological Evaluations of Torture*. Boston, MA: Physicians for Human Rights; 2001.
11. Rasmussen OV. Medical aspects of torture. *Dan Med Bull*1990;37(suppl 1):1–88.
12. Sinding R, Smidt-Nielsen. The late ear, nose, and throat region sequelae of torture. *Torture* 1999;9:20–22.
13. Asgary RG, Metalios EE, Smith CL, Paccione GA. Evaluating asylum seekers/torture survivors in urban primary care: a collaborative approach at the Bronx Human Rights Clinic. *Health Hum Rights* 2006;9:164–179.
14. Physical health: types of torture. Boston Center for Refugee Health and Human Rights Web site. Available at: http://www.bcrhhr.org/pro/course/physical_health_types_of_torture.html. Accessed July 29, 2009.
15. Moisander PA, Edston E. Torture and its sequel—a comparison between victims from six countries. *Forensic Sci Int* 2003;137:133–140.
16. Stanley RB, Hanson DG. Manual strangulation injuries of the larynx. *Arch Otolaryngol* 1983;109:344–347.
17. Moreno A, Grodin MA. Torture and its neurological sequelae. *Spinal Cord* 2002;40:213–223.
18. Vogel H, Schmitz-Engels F, Grillo C. Radiology of torture. *Eur J Radiol* 2007;63:187–204.
19. Mirzaei S, Knoll P, Lipp R, et al. Bone scintigraphy in screening of torture survivors. *Lancet* 1998;352:949–951.
20. Weinstein HM, Dansky L, Iacopino V. Torture and war trauma survivors in primary care practice. *West J Med* 1996;165:112–118.
21. National Capacity Building Project. Center for Victims of Torture Web site. Available at: http://www.cvt.org/files/pg100/Healing_the_Hurt_Ch5.pdf. Accessed July 29, 2009.
22. Fazel M, Wheeler J, Danesh J. Prevalence of serious mental disorder in 7000 refugees resettled in western countries: a systematic review. *Lancet* 2005;365:1309–1314.
23. Crosby SS, Mashour GA, Grodin MA, Jiang Y, Osterman J. Emergence flashback in a patient with posttraumatic stress disorder. *Gen Hosp Psychiatry* 2007;29:169–171.
24. National Consortium of Torture Treatment Programs (NCTTP) Web site. Available at: http://ncttp.dataweb.com/default.view. Accessed July 29, 2009.
25. Boston Center for Refugee Health and Human Rights Web site. Available at: http://www.bu.edu/bcrhhr. Accessed July 29, 2009.

PREVENTING CHRONIC DISEASE
PUBLIC HEALTH RESEARCH, PRACTICE, AND POLICY
VOLUME 7: NO. 3 MAY 2010

ORIGINAL RESEARCH

Chronic Disease and Its Risk Factors Among Refugees and Asylees in Massachusetts, 2001-2005

Nameeta M. Dookeran, MD, MSc; Tracy Battaglia, MD, MPH; Jennifer Cochran, MPH; Paul L. Geltman, MD, MPH

Suggested citation for this article: Dookeran NM, Battaglia T, Cochran J, Geltman PL. Chronic disease and its risk factors among refugees and asylees in Massachusetts, 2001-2005. Prev Chronic Dis 2010;7(3). http://www.cdc.gov/pcd/issues/2010/may/09_0046.htm. Accessed [*date*].

PEER REVIEWED

Abstract

Introduction

Better understanding of the health problems of refugees and people who are granted political asylum (asylees) in the United States may facilitate successful resettlement. We examined the prevalence of risk factors for and diagnoses of chronic disease among these groups in Massachusetts.

Methods

We retrospectively analyzed health screening data from 4,239 adult refugees and asylees who arrived in Massachusetts from January 1, 2001, through December 31, 2005. We determined prevalence of obesity/overweight, hypertension, coronary artery disease (CAD), diabetes, and anemia. Analyses included multivariate logistic regression to determine associations between CAD and diabetes with region of origin.

Results

Almost half of our sample (46.8%) was obese/overweight, and 22.6% had hypertension. CAD, diabetes, and anemia were documented in 3.7%, 3.1%, and 12.8%, respectively. People from the Europe and Central Asia region were more likely than those from other regions to have CAD (odds ratio, 5.55; 95% confidence interval, 2.95-10.47).

Conclusions

The prevalence of obesity/overweight and hypertension was high among refugees and asylees, but the prevalence of documented CAD and diabetes was low. We noted significant regional variations in prevalence of risk factors and chronic diseases. Future populations resettling in the United States should be linked to more resources to address their long-term health care needs and to receive culturally appropriate counseling on risk reduction.

Introduction

The United States has a longstanding humanitarian commitment to the resettlement of refugees from overseas. Each year, the number of people granted refugee and political-asylum status in the United States fluctuates based on variations in the stability of other countries, the global political climate, and domestic resettlement targets. The largest number of refugee admissions in 2008 was to the United States (68% of the 80,800 resettled refugees worldwide), but Australia, Canada, and Sweden had higher per capita admission rates (52.4, 32.5, and 24.3 refugees per 100,000 residents, respectively) than the United States (19.8 refugees per 100,000 residents) (1-3). In the 3 years from 2006 through 2008, the United States approved an average of 24,750 claims for asylum per year; the 3 leading countries of nationality (China, Colombia, and Haiti) together constituted 36.4% to 44.0% of all approvals (4).

Refugees and asylees (people who are granted asylum) are people outside of their country of origin who are unable

or unwilling to return to that country because they have experienced, or have a legitimate fear of, persecution on the grounds of race, religion, nationality, membership in a particular social group, or political affiliation (5). People who are granted refugee status and admission to the United States apply while overseas after having fled their home country or, for certain nationalities, while in-country. In contrast, people who seek political asylum do so either some time after entry into the United States or on arrival at a US port of entry. Historically, asylum applicants were in the United States for many years before being granted asylum because of delays in filing and processing asylum applications. Recently, this difference in time in the United States between refugees and asylees has lessened, in part because 1995 federal immigration legislation required potential asylees to file asylum applications within 1 year of arrival (6).

Because refugees and asylees differ in how long they have been in the United States, their countries of origin, and their socioeconomic circumstances, they likely have different health care needs. The Massachusetts Refugee Health Assessment Program (RHAP), a partnership between the Department of Public Health and contracted private, mostly federally qualified clinics, was established in 1995 to perform health screenings of refugees and other people who were eligible for refugee benefits. High rates of CAD in Russia include asylees, Cuban and Haitian entrants, certain Amerasians (mostly from Vietnam), and victims of human trafficking (7). Asylees were effectively denied access to RHAP services until 2000, when the starting date of time-limited eligibility for services was changed from the date of physical entry into the United States to the later date of asylum approval (8).

Domestic refugee health assessment programs, such as RHAP, have traditionally focused on identification and treatment of infectious diseases, although such programs also serve as a bridge to primary care. Few studies have focused on the screening of newly arrived refugees in the United States for chronic diseases, mental illness, or substance abuse, despite their relevance in these populations (9-14). Asylees may also be at risk of developing chronic diseases through acculturation while living as marginalized residents of low-income, urban neighborhoods in the United States before being granted asylum status.

The prevalence of chronic disease is high in many of the countries where refugees and asylees live before resettling in the United States. World Health Organization data show higher chronic-disease–related death rates in low- and middle-income countries compared with Canada or the United Kingdom (15). During the past 15 years, the largest group of refugees entering the United States has been from the nations that were formed from the former Soviet Union. Among this group, the Russian Federation in particular has seen growing mortality from preventable causes other than communicable disease, and cardiovascular disease is the leading cause of death (16). In the wake of the collapse of the Soviet Union, Russian life expectancy has declined as rates of nutritional deficiency and alcoholism have risen (17-19).

The changing demographics of both refugees and asylees entering the US health care system may result in greater health care needs for chronic, noninfectious diseases. However, programs designed to assess refugee health care needs are not generally structured to address chronic health problems. The objectives of this study were to determine the documented prevalence of risk factors for, and diagnoses of, chronic diseases among refugees and asylees who received RHAP health screening and to determine whether differences in prevalence of chronic disease and risk factors were associated with region of origin or visa category.

Methods

We performed a retrospective cross-sectional study using RHAP data from health screenings of asylees and refugees. For the purposes of this article, the term "refugees" includes people newly arrived in the United States from overseas (ie, true refugees), derivative asylees (ie, people arriving from overseas to reunite with immediate family members previously granted asylum in the United States), and Cuban, Haitian, and Amerasian special entrants. Eligible participants were aged 18 years or older, had entered the United States from January 1, 2001, through December 31, 2005, and had completed the RHAP screening (7). The institutional review board of Boston University Medical Center approved and monitored the conduct of this study, and the Massachusetts Department of Public Health approved the public release of this data analysis.

In the RHAP electronic database, the Massachusetts Department of Public Health maintains clinical and public health data on asylees and refugees, derived from official

government arrival notifications and RHAP reporting forms submitted by contracted health assessment clinical sites. Government arrival notifications are the source of basic demographic information (eg, patient age, sex, country of origin) and, in the case of refugees, medical diagnoses documented in reports from medical examinations performed overseas before arrival in the United States. RHAP reporting forms are the source of additional medical diagnoses and information obtained during refugee and asylee screening in the United States; they comprise a history and physical examination, immunizations, and a set of standard (eg, stool ova and parasites, complete blood counts, urinalyses) and optional tests based on individual health needs.

Risk factors for chronic disease included evidence of obesity (body mass index [BMI], ≥30 kg/m^2) or being overweight (BMI, 25.0-29.9 kg/m^2) and provider documentation of hypertension (including people with a single high blood pressure [systolic blood pressure ≥140 mm Hg] measurement) (20-22). Not all people with 1 elevated systolic blood pressure reading have true hypertension, but they require clinical follow-up because of their risk of hypertension. Chronic disease measures available for this study included provider documentation of CAD and diabetes (including evidence of glucosuria on urinalysis), and evidence of anemia (by hemoglobin values of <13 g/dL in men and <12 g/dL in women) (23).

In describing the population that used services, we first determined the number of refugees seen in RHAP from 2001 through 2005 by year of US entry and the number of asylees seen by year in which status was granted. We then described all people who completed RHAP screening by sex, visa category, age, and region of origin. The 5 regions of origin represented 92 countries.

For our main analyses, we determined the prevalence of obesity/overweight, hypertension, CAD, diabetes, and anemia, overall and by region of origin. We also determined the prevalence of obesity/overweight by age group. We used SAS version 9.1 (SAS Institute, Inc, Cary, North Carolina) to conduct multivariate logistic regression to examine associations of CAD and diabetes with being from the Europe and Central Asia region (including countries of the former Soviet Union and the former Yugloslavia), adjusting for age, sex, and BMI as covariates in the model. Visa category was not included in regression models because of the low numbers of asylees in the overall population and concerns about covariation of visa category with the more robust place-of-origin variable. Among refugees only, we also examined the proportion of documented diagnoses of CAD and diabetes that originated in reports from overseas medical examinations performed before US arrival and participation in RHAP screening.

Results

Of the 5,141 adult refugees and asylees with dates of entry from 2001 through 2005, RHAP documentation was available for 4,239 (82.5%) who completed health screening. Those who completed RHAP screening were similar to those who did not with respect to mean age (37.7 vs 36.5 years) and sex (49.8% vs 52.4% women). They differed in respect to country of origin (43.8% of completers vs 25.3% of noncompleters were from Europe and Central Asia) and visa category (11.2% of completers vs 13.5% of noncompleters were asylees).

The distribution of visa categories among people who received RHAP services varied by year of eligibility (Figure) and reflect the US allocations of visas each year (24). The reduced numbers of refugees from 2001 through 2005 reflect the government's limited processing of visa applications of refugees overseas after the September 11 terrorist attacks. The top 2 regions of origin of all people who completed RHAP screening were Europe and Central Asia and Africa (Table 1). Among the 3,765 refugees and the 474 asylees who completed RHAP screening, the top regions of origin respectively were Europe and Central Asia (47.8%) and Africa (52.7%). Compared with asylees, refugees had a higher mean (SD) age (38.8 [16.0] years vs 34.8 [10.5] years for asylees) and a slightly lower proportion of women (50.3% vs 52.7% of asylees).

We found differences in sex, visa category, and mean age by region of origin (Table 1). Women accounted for approximately half of asylees/refugees from all 5 regions. Asylees accounted for as little as 3.1% of people from Europe and Central Asia and as much as 28.5% of those from Latin America and the Caribbean. The mean age was highest for people from Europe and Central Asia and lowest for those from Africa.

Overall, almost one-fifth of this sample was obese, and more than one-fourth was overweight (Table 2). The largest proportions of obese and overweight people were

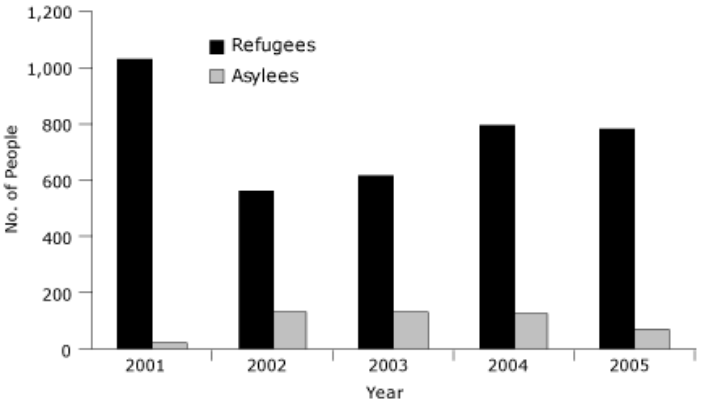


Figure. Distribution of refugees and asylees who received health assessment services in Massachusetts, 2001-2005 (N = 4,239).

from Europe and Central Asia. Among 18- to 49-year-olds, more than one-fourth were overweight and 15.9% were obese. Among 50- to 79-year-olds, 34.3% were overweight and 31.5% were obese. Almost one-fourth had hypertension diagnoses, again with documentation highest among people from Europe and Central Asia. Those from East and Southeast Asia had the lowest prevalence of obesity/overweight (3.6% and 21%, respectively) and the lowest prevalence of hypertension.

Documented chronic diseases varied by region of origin (Table 2). People from Europe and Central Asia contributed disproportionately to documented diagnoses of coronary artery disease (CAD). Anemia was highest among people from Africa and lowest among those from East and Southeast Asia.

In logistic regression models adjusting for age, sex, and BMI, people from Europe and Central Asia were significantly more likely than others to have CAD (adjusted odds ratio [AOR], 5.55; 95% confidence interval [CI], 2.95-10.47). Additionally, they were slightly less likely to have diabetes (AOR, 0.74; 95% CI, 0.49-1.13), but this latter finding was not significant.

Among the total of 157 diagnoses of CAD, 153 were among refugees rather than asylees. Most (81%) of these 153 refugee diagnoses had been entered in the RHAP database from overseas medical examination reports rather than from new findings during RHAP screening. Most (95%) of these 153 refugee diagnoses were among people from Europe and Central Asia. Among the total of 131 diagnoses of diabetes, 71 were among refugees. Almost half (49%) of these 71 diagnoses had been entered in the RHAP database from overseas medical examination reports. As with CAD, most (61%) of the 71 refugee diagnoses were among people from Europe and Central Asia.

Discussion

Region of origin was strongly associated with prevalence of risk factors for and presence of the chronic diseases assessed in this study, with the exception of diabetes. Associations with visa category were less consistent; however, because of their high concentration among people from Europe and Central Asia, refugees were significantly more likely than asylees to have certain risk factors or chronic diseases, particularly CAD. We found that almost one-fifth of our sample were obese, more than one-fourth were overweight, and almost one-fourth had hypertension. In comparison, the overall rates of documented CAD and diabetes were low. Refugees and asylees from the Europe and Central Asia region had the highest prevalence of obesity/overweight and hypertension and were more than 5 times more likely to have documented CAD compared with those from other regions. Regional differences in anemia prevalence in this young study sample were also apparent, suggesting other underlying chronic disease or nutritional deficiencies that varied by region.

Few studies of chronic disease among United States refugee populations exist, necessitating comparison of our findings with those of studies of immigrants as well as refugees. A recent study of 459 refugee psychiatric patients found the prevalence of hypertension and diabetes to be 42.0% and 15.5%, respectively (25). This was significantly higher than US norms and was especially pronounced in people younger than 65 years. Rates of obesity were also high, especially among Bosnians (54.5%), similar to our findings among people from Europe and Central Asia. In another study of Russian-speaking adult immigrants in New York, 53.8% had hypertension and 33.2% were obese, significantly higher prevalence rates than among other non-Hispanic whites after age adjustment (26). Lastly, in a nationally representative study of 6,421 adult immigrants with newly acquired legal permanent residence, the adjusted prevalence of obesity/overweight ranged from 36.5% to 65.9% for men and from 21.7% to 53.3% for women across all regions (27). The prevalence was lowest among men and women from Asia (similar to our study findings) and highest among men from the Latin America

and Caribbean region and women from the Middle East and North Africa region. Higher prevalence of risk factors and chronic diseases found in the studies above may be related to more acculturation to US lifestyle (28,29).

The low prevalence of CAD and diabetes found in our study may be accurate in this young population of primarily recently arrived refugees. It may also indicate inadequate time or resources for diagnosis of disease during either overseas or US health screening. Despite the overall low prevalence of CAD, the significantly increased likelihood of CAD among people from Europe and Central Asia compared with those from other regions may reflect the high burden of this disease in Russia, where cardiovascular disease is the leading cause of death (16). The latter may be due in part to high rates of smoking and hypertension in this region. One unexpected finding was the lower (but not significant) likelihood of diabetes in people from Europe and Central Asia compared with all other regions. This could be related to distinct differences in dietary patterns in Europe and Central Asia, including an increase in moderate alcohol consumption, which has been postulated in a meta-analysis of epidemiologic data on diabetes risk factors to reduce risk for development of type 2 diabetes (30).

One of this study's main strengths was the large sample size and demographic diversity of the refugees and asylees in Massachusetts. The large numbers of refugees and asylees in the RHAP database facilitated comparisons of the prevalence of risk factors and diagnoses of chronic diseases across regions of origin that could not have been done using a sample drawn from a single clinic. These comparisons are likely generalizable to other refugees and asylees resettling across the United States during the study period. However, they may be less generalizable to refugee/asylee populations entering the United States in other years because the regions of origin represented, as well as the diet and lifestyle patterns in a given region, may change over time.

The data available from the RHAP database were somewhat limited. Although refugees in the RHAP are typically seen within 90 days of arrival in the United States, it is likely that asylees had been in the United States for a longer time before RHAP screening, thus increasing chances of acculturation to US diet and lifestyle (7). However, data were not available to quantify these times more precisely. In addition, CAD, diabetes, and hypertension may have been underreported because these diagnoses were based on provider documentation either from overseas medical examinations or domestic health screening. On the other hand, we were able to extract more objective measures from the RHAP database to quantify obesity/overweight, elevated blood pressure, glucosuria, and anemia.

In summary, although rates of CAD and diabetes were low, this study found a high prevalence of risk factors for chronic disease such as obesity/overweight and hypertension. Findings suggest that refugees and asylees from Europe and Central Asia fall into a high-risk category. Future populations resettling in the United States should be linked to more resources to address their long-term health care needs and to receive culturally appropriate counseling on risk reduction. Further studies may shed more light on differences in risk among different subpopulations of refugees and asylees, but more programs are needed to help establish primary care after domestic health screening. Primary care will increase the overall health of these populations and the likelihood that they will be able to successfully integrate into United States society over time.

Acknowledgments

This research was supported in part by the Massachusetts Department of Public Health and the General Internal Medicine Fellowship, Boston University School of Medicine. We gratefully acknowledge the statistical support provided by Thinh Nguyen. A portion of this research was presented as a poster at the 2007 New England Regional Society of General Internal Medicine Conference.

Author Information

Corresponding Author: Paul L. Geltman, MD, MPH, Refugee and Immigrant Health Program, Massachusetts Department of Public Health, Department of Pediatrics, 305 South St, Jamaica Plain, MA 02130. Telephone: 617-983-6593. E-mail: paul.geltman@state.ma.us. Dr Geltman is also affiliated with Boston University School of Medicine.

Author Affiliations: Nameeta M. Dookeran, Brigham and Women's Hospital, Boston, Massachusetts; Tracy Battaglia, Boston University School of Medicine, Boston,

Massachusetts; Jennifer Cochran, Massachusetts Department of Public Health, Boston, Massachusetts.

References

1. 2008 Global trends: refugees, asylum-seekers, returnees, internally displaced and stateless persons. United Nations High Commissioner for Refugees; 2009. http://www.unhcr.org/4a375c426.html. Accessed January 14, 2010.
2. 2008 Population statistics. CIA world factbook. https://www.cia.gov/library/publications/the-world-factbook/. Accessed July 7, 2009.
3. State and county QuickFacts. US Census Bureau. http://quickfacts.census.gov/qfd/states/00000.html. Accessed January 19, 2009.
4. Martin DC, Hoefer M. Refugees and asylees: 2008. US Department of Homeland Security, Office of Immigration Statistics; 2009. http://www.dhs.gov/xlibrary/assets/statistics/publications/ois_rfa_fr_2008.pdf. Accessed July 16, 2009.
5. UN Refugee Convention: Convention Relating to the Status of Refugees. In: Twenty-four human rights documents. New York (NY): Columbia University Center for the Study of Human Rights; 1992.
6. 1996 Illegal Immigration Reform and Immigrant Responsibility Act. PL 104-208; Title VI; Sec 604.
7. Geltman PL, Cochran J. A private-sector preferred provider network model for public health screening of newly resettled refugees. Am J Public Health 2005;95(2):196-9.
8. State Letter #00-12: Asylee eligibility for refugee resettlement program benefits. US Department of Health and Human Services, Administration for Children and Families. http://www.acf.hhs.gov/programs/orr/policy/sl00-12.htm. Accessed January 14, 2010.
9. Ackerman LK. Health problems of refugees. J Am Board Fam Pract 1997;10(5):337-48.
10. Adams KM, Gardiner LD, Assefi N. Healthcare challenges from the developing world: post-immigration refugee medicine. BMJ 2004;328(7455):1548-52.
11. Power DV, Shandy DJ. Sudanese refugees in a Minnesota family practice clinic. Fam Med 1998;30(3):185-9.
12. Fritz MJ, Hedemark LL. Somali refugee health screening in Hennepin County. Minn Med 1998;81(4):43-7.
13. Chapman DP, Perry GS, Strine TW. The vital link between chronic disease and depressive disorders. Prev Chronic Dis 2005;2(1). http://www.cdc.gov/pcd/issues/2005/jan/04_0066.htm. Accessed January 14, 2010.
14. Nelson KR, Bui H, Samet JH. Screening in special populations: a "case study" of recent Vietnamese immigrants. Am J Med 1997;102(5):435-40.
15. Strong K, Mathers C, Leeder S, Beaglehole R. Preventing chronic diseases: how many lives can we save? Lancet 2005;366(9496):1578-82.
16. Levintova M. Cardiovascular disease prevention in Russia: challenges and opportunities. Public Health 2006;120(7):664-70.
17. Notzon FC, Komarov YM, Ermakov SP, Sempos CT, Marks JS, Sempos EV. Causes of declining life expectancy in Russia. JAMA 1998;279(10):793-800.
18. Kohlmeier L, Mendez M, Shalnova S, Martinchik A, Chakraborty H, Kohlmeier M. Deficient dietary iron intakes among women and children in Russia: evidence from the Russian Longitudinal Monitoring Survey. Am J Public Health 1998;88(4):576-80.
19. Rahav G, Hasin D, Paykin A. Drinking patterns of recent Russian immigrants and other Israelis: 1995 national survey results. Am J Public Health 1999;89(8):1212-6.
20. Clinical guidelines on the identification, evaluation, and treatment of overweight and obesity in adults — the evidence report. National Institutes of Health. Obes Res 1998;6(Suppl 2):51-209S.
21. Obesity: preventing and managing the global epidemic. Report of a WHO consultation. World Health Organ Tech Rep Ser 2000;894:1-253.
22. Chobanian AV, Bakris GL, Black HR, Cushman WC, Green LA, Izzo JL Jr, 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 2003;289(19):2560-72.
23. Blanc B, Finch CA, Hallberg L. Nutritional anaemias. Report of a WHO scientific group. World Health Organ Tech Rep Ser 1968;405:1-40.
24. Proposed refugee admissions for fiscal year 2009. Report to the Congress. http://www.state.gov/documents/organization/113507.pdf. Accessed January 14, 2010.
25. Kinzie J, Riley C, McFarland B, Hayes M, Boehnlein J, Leung P, et al. High prevalence rates of diabetes and hypertension among refugee psychiatric patients. J Nerv Ment Dis 2008;196:108-12.
26. Hosler AS, Melnk TA, Spence MM. Diabetes and its related risk factors among Russian-speaking immi-

grants in New York State. Ethn Dis 2004;14:372-7.
27. Roshania R, Narayan KMV, Oza-Frank R. Age at arrival and risk of obesity among US immigrants. Obesity 2008;16:2669-75.
28. Kandula NR, Diez-Roux AV, Chan C, Daviglus ML, Jackson SA, Ni H, et al. Association of acculturation levels and prevalence of diabetes in the Multi-Ethnic Study of Atherosclerosis (MESA). Diabetes Care 2008;31:1621-8.
29. Gomez SL, Kelsey JL, Glaser SL, Lee MM, Sidney S. Immigration and acculturation in relation to health and health-related risk factors among specific Asian subgroups in a health maintenance organization. Am J Public Health 2004;94(11):1977-84.
30. Carlsson S, Hammar N, Grill V. Alcohol consumption and type 2 diabetes meta-analysis of epidemiological studies indicates u-shaped relationship. Diabetologia 2005;48(6):1051-4.

Tables

Table 1. Most Commonly Represented Countries/Areas Within the 5 Regions of Origin of People Who Received Refugee Health Assessment Services, Massachusetts, 2001-2005 (N = 4,239)

Countries/Areas	All, n (% Total)	Women, n (% Region)	Asylees, n (% Region)	Age, Mean (SD), y
Europe and Central Asia	1,858 (43.8)	980 (52.7)	57 (3.1)	43.8 (17.5)
Former Soviet Union	1,634 (38.5)	871 (53.3)	32 (2.0)	44.8 (17.8)
Former Yugoslavia	195 (4.6)	94 (48.2)	1 (0.5)	35.7 (12.9)
Albania	29 (0.7)	15 (51.7)	24 (82.8)	37.4 (11.5)
Africa	1,497 (35.3)	704 (47.0)	250 (16.7)	31.8 (13.1)
Somalia	493 (11.6)	242 (49.1)	19 (3.9)	34.3 (14.8)
Liberia	305 (7.2)	176 (57.7)	22 (7.2)	31.9 (13.6)
Sudan	220 (5.2)	39 (17.7)	7 (3.2)	25.5 (7.7)
East and Southeast Asia	338 (8.0)	164 (48.5)	55 (16.3)	36.3 (11.6)
Vietnam	185 (4.4)	91 (49.2)	0 (0.0)	35.6 (10.0)
Cambodia	99 (2.3)	53 (53.5)	49 (49.5)	36.9 (13.4)
Burma	24 (0.6)	5 (20.8)	1 (4.2)	36.3 (10.0)
Near East and South Asia	213 (5.0)	107 (50.2)	17 (8.0)	35.0 (12.8)
Afghanistan	135 (3.2)	78 (57.8)	0	36.2 (13.5)
Iran	44 (1.0)	14 (31.8)	7 (15.9)	32.4 (12.4)
Iraq	25 (0.6)	11 (44.0)	3 (12.0)	34.0 (8.0)
Latin America and Caribbean	333 (7.9)	154 (46.3)	95 (28.5)	33.2 (9.6)
Haiti	233 (5.5)	97 (41.6)	60 (25.8)	31.8 (7.7)
Cuba	44 (1.0)	19 (43.2)	0	38.9 (12.4)
Colombia	42 (1.0)	30 (71.4)	27 (64.3)	36.6 (11.5)
All regions	4,239 (100.0)	2,109 (49.8)	474 (11.2)	37.7 (15.8)

Table 2. Medical Conditions by Region of Origin of People Who Received Refugee Health Assessment Services, Massachusetts, 2001-2005 (N = 4,239)[a]

Region	Risk Factors, n (% Region)			Chronic Diseases, n (% Region)		
	Obesity[b]	Overweight[b]	HTN[c]	CAD	Diabetes[d]	Anemia[e]
Europe and Central Asia	508 (27.3)	580 (31.2)	599 (32.2)	145 (7.8)	65 (3.5)	176 (9.5)
Africa	199 (13.3)	362 (24.2)	245 (16.4)	8 (0.5)	37 (2.5)	294 (19.6)
East and Southeast Asia	12 (3.6)	71 (21.0)	33 (9.8)	2 (0.6)	12 (3.6)	18 (5.3)
Near East and South Asia	29 (13.6)	58 (27.2)	25 (11.7)	1 (0.5)	6 (2.8)	26 (12.2)
Latin America and Caribbean	62 (18.6)	104 (31.2)	58 (17.4)	1 (0.3)	11 (3.3)	30 (9.0)
All regions	810 (19.1)	1,175 (27.7)	960 (22.6)	157 (3.7)	131 (3.1)	544 (12.8)

Abbreviations: HTN, hypertension; CAD, coronary artery disease.
[a] χ^2 Statistical testing was used to determine association between having a given chronic disease or risk factor and region of origin: $P = .51$ for diabetes, $P < .001$ for all other conditions.
[b] Obesity defined as body mass index ≥30 kg/m^2, overweight defined as 25.0-29.9 kg/m^2.
[c] Defined as diagnosis of HTN or measurement of systolic blood pressure ≥140 mm Hg.
[d] Included presence of glucose on urinalysis.
[e] Hemoglobin <13 g/dL or hematocrit <41% (men) and hemoglobin <12 g/dL or hematocrit <36% (women).

INNOVATION AND IMPROVEMENT

Tracking Abnormal Cervical Cancer Screening: Evaluation of an EMR-Based Intervention

Elizabeth A. Dupuis, MD[1], Hilary F. White[1], Daniel Newman, MD[2], Jerome E. Sobieraj, MD[2], Manjusha Gokhale, MA[3], and Karen M. Freund, MD, MPH[1]

[1]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; [2]Section of General Internal Medicine, Boston University School of Medicine, Boston, MA, USA; [3]Data Coordinating Center, Boston University School of Public Health, Boston, MA, USA.

INTRODUCTION: System level barriers have been associated with inadequate follow-up of abnormal cervical cytology.

OBJECTIVE: The aim of this study was to develop and evaluate an electronic tracking system to improve follow-up of abnormal Pap tests.

PROGRAM DESCRIPTION: We implemented an electronic medical record (EMR)-based Pap test tracking system at two clinical practices at an inner-city academic health center. The system generated a provider-specific monthly report of all abnormal Pap results, and provided a patient-specific Pap tracking table embedded in the EMR for each subject.

EVALUATION: We compared abnormal Pap test follow-up rates for the 24 months pre-intervention with rates 12 months following its implementation (post-intervention). The evaluation followed all subjects for 12 months from the date of their abnormal Pap test, looking for diagnostic resolution.

RESULTS: Subjects were young women (mean age = 30.5) of primarily white (42%) and African American (37%) descent, who spoke English (88%). Forty-eight percent were insured through publicly subsidized insurance. Controlling for type of abnormality and practice location, the adjusted mean time to resolution decreased significantly from 108 days (confidence interval, CI 105–112 days) in the pre-intervention period to 86 days (CI 81–91 days).

CONCLUSION: Our study cannot demonstrate that with follow up, we directly avoided cases of invasive cervical cancer. However, we show that in an at-risk urban population, an automated, EMR-based tracking system reduced the time to resolution, and increased the number of women who achieved diagnostic resolution.

KEY WORDS: pap smear; cervical cancer; papillomavirus.
J Gen Intern Med 25(6):575–80
DOI: 10.1007/s11606-010-1287-z

Received June 19, 2009
Revised December 10, 2009
Accepted December 29, 2009
Published online March 4, 2010

INTRODUCTION

Screening for cervical cancer with a Pap test is only as successful as the follow-up rate for an abnormal result. If a patient has a Pap test, yet does not receive appropriate follow-up for an abnormal result, then the opportunity to prevent or treat pre cancerous lesions or cervical cancer is missed and the Pap test is ineffective. This is specifically an issue in lower income and minority populations who experience a higher risk of cervical cancer[1,2] and a higher rate of inadequate abnormal Pap test follow-up[3]. Multiple studies have documented the problem of inadequate follow-up of abnormal Pap tests[3–11].

AIM

With the advent of electronic medical record (EMR) systems, there is great potential to address inadequate follow-up from a systems point of view. We developed a tracking system for our internal EMR, and evaluated this tracking system as an intervention to improve adequate follow-up of abnormal Pap tests.

PROGRAM DESCRIPTION

We developed a tracking system that has two components:

1) A tracking report of abnormal Pap tests generated for the providers each month
2) A Pap test tracking table embedded in the EMR

The first component, the tracking report, identified potential cases by searching for Pap test orders placed by a provider in the EMR. An interface between the EMR and the pathology reports was developed and from these text fields created a document in the EMR. It was then possible to extract and track the cytology report corresponding to the day the order was placed. Pathology reports are provided in relatively standard text-only format which is scanned for specific phrases. Initially, work was done in coordination with the pathology department to determine standardized Pap test result language. In addition, all possible combinations of added spaces, added hyphens, and lower-case versus capitalized letters were accounted for so that abnormal test result would not be excluded. This process involved multiple iterations of comparing tracking reports directly from pathology, and adjusting the text parse filters to ensure capture of any missed results. Additionally, a hierarchy of abnormality severity was developed

Table 1. Field Provided in Monthly Provider Pap Test Tracking Reports

Demographic information:
• Practice location
• Provider name
• Subject medical record number (MRN)
• Subject name
Pap test information:
• Date of Pap test
• Result of Pap test
• HPV status (positive or negative, if tested)
• Result of LAST abnormal Pap test (if any)
• Date of LAST abnormal Pap test (if any)
Follow-up information:
• Date subject contacted of an abnormal result
• Method used to contact subject (eg. letter, phone call)
• Date of GYN follow-up appointment
• Status of GYN follow-up appointment (eg. future, arrived)
• Number of cancelled GYN appointments
• Practice location of GYN follow-up appointment

so that if a Pap result mentioned both ASCUS (atypical squamous cells of undetermined significance) and HGSIL (high grade squamous intraepithelial lesion), for example, then it would be labeled with the more severe abnormality (HGSIL).

Documented phone and letter contacts from standardized templates in the EMR were included with dates. Appointment data for this tracking report were extracted from another system that manages the outpatient clinic scheduling. Colposcopy appointment dates, location and status appeared on the report, including cancelled and no-show appointments. When a gynecologist performs a colposcopy, they use a standard colposcopy procedure template, which can therefore be tracked. A completed colposcopy was considered resolution of the screening Pap abnormality, and at this point the subject fell off the tracking report. The tracking reports were generated in spreadsheet format with the relevant provider and subject identifiers and results (Table 1). The report was cumulative, meaning that unresolved abnormal Pap test results remained on the report until resolved.

The second component of the tracking system was an individual Pap test tracking table within any individual subject record (see Fig. 1). This table showed the details of all past Pap test results, linked patient contacts, appointments and gynecology pathology results. The EMR tracking table gave providers efficient access to current Pap test status when seeing a patient in the office, and was another point of intervention during this visit.

The tracking reports were distributed to each provider monthly and included all of their patients who had had an unresolved Pap test abnormality (Fig. 2). We purposefully delayed the reporting of the abnormal Pap test results by one to two months to allow time for the subject to be contacted and for the colposcopy to be scheduled, and for the list to reflect patients with true delays. The standard manner for informing providers about their abnormal Pap tests did not change, where the Pap test result arrived directly in the ordering provider's EMR inbox for their review and action. Therefore, the paper tracking report serves as a second notice to providers of any abnormal Pap results, and highlighted patients who either did not keep or did not schedule a gynecology follow-up appointment, or who may never have

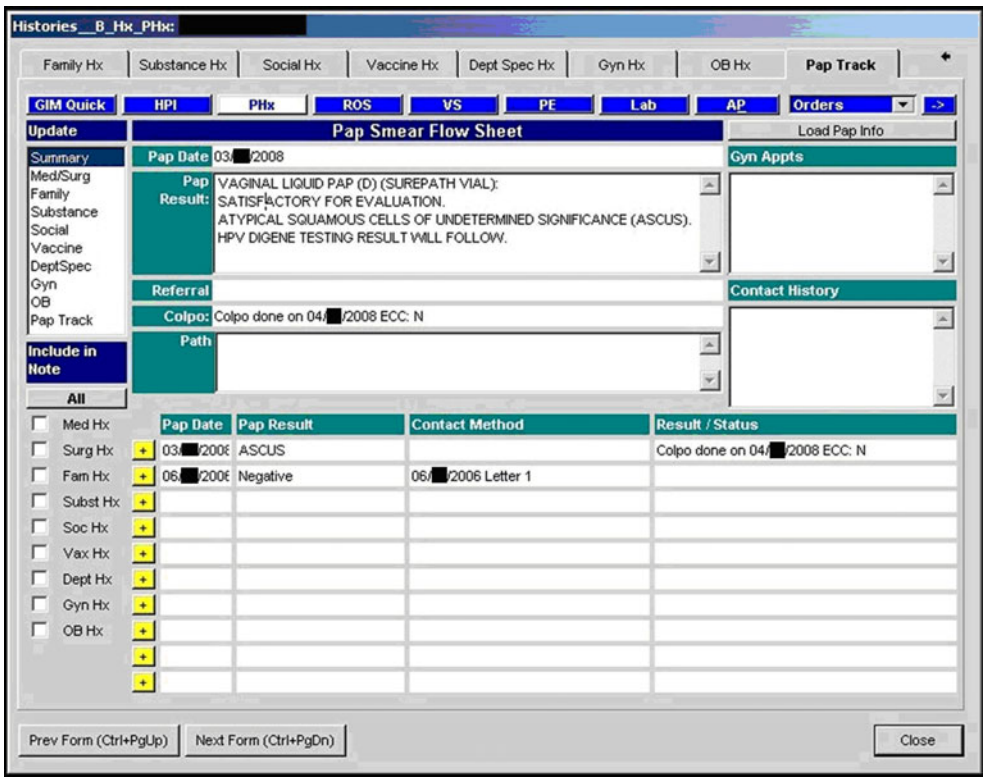


Figure 1. Example of an EMR Pap tracking table.

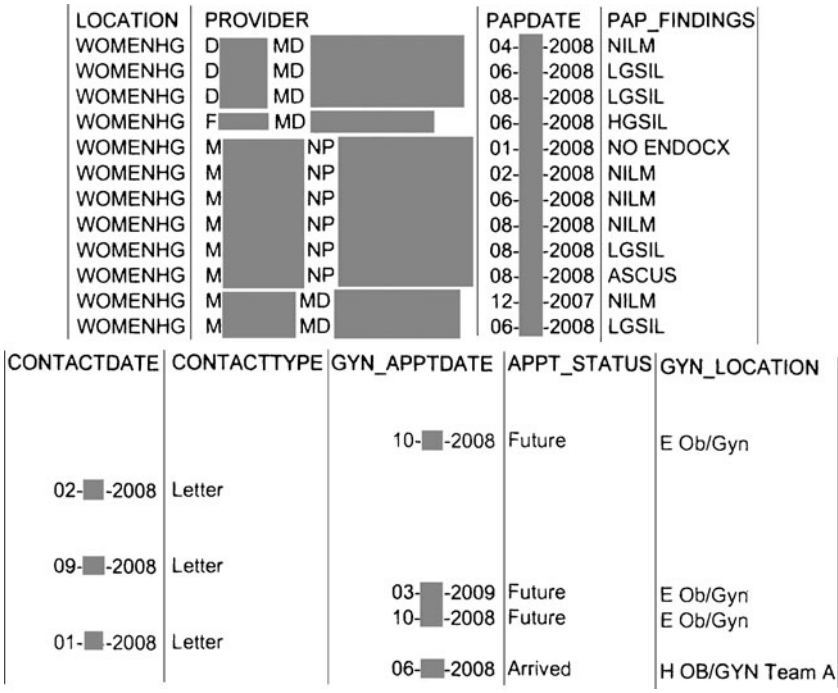


Figure 2. Example of a monthly abnormal Pap report.

been notified of their abnormal test results. Once the providers received copies of their individual Pap tracking reports there was no specific protocol about how they managed the information on these reports. No additional resources were given to providers to manage the information on the tracking reports.

EVALUATION

Study Design

We used a pre-test/post-test study design to evaluate whether implementation of a Pap tracking system 1) reduced the number of subjects with inadequate follow-up of abnormal Pap tests, and 2) reduced the time to follow-up. We compared inadequate Pap test follow-up rates prior to availability of the tracking system (pre-intervention) with inadequate follow-up rates following its implementation (post-intervention) at two clinical practices at an inner-city academic health center. The pre-intervention time period was December 2004 to December 2006. We allowed a 3-month implementation period for the tracking system, during which providers were trained and supported in its use. The tracking system was formally implemented on February 1, 2007 at one site, and April 1, 2007 at the second site. The post-intervention time period was therefore March 2007 to April 2008 for the first site, and May 2007 to June 2008 for the second site. All test results were followed for 12 months to determine if a diagnostic evaluation had been completed.

Study Subjects

Eligible subjects were 18 years of age or older, and had one of the following abnormal Pap test results:

1) Atypical squamous cells of undetermined significance with positive high-risk HPV serotype (ASCUS/HPV+)
2) Low-grade squamous intraepithelial lesion (LGSIL)
3) Atypical glandular cells of undetermined significance (AGC/AGUS)
4) Atypical squamous cells: cannot exclude high-grade squamous intraepithelial lesion (ASC-H)
5) High-grade squamous intraepithelial lesion (HGSIL)
6) Carcinoma in situ, or invasive cancer

Data Collection

We obtained an independent list from the pathology department of all abnormal Pap tests during the study time periods for each provider in the two clinical practices. Data were collected through retrospective electronic chart abstraction of all women with abnormal Pap tests, reviewed for 12 months following the abnormal test.

Independent Variables

Race/ethnicity was documented in the EMR as a single set of seven mutually exclusive responses, which we collapsed into "white," "black/African American" and "other." Primary language (nine categories) was collapsed into "English" and "non-English." Health insurance coverage was grouped into private, public and no insurance. The type of cervical abnormality was collapsed into ASCUS/HPV+, AGC/AGUS, LGSIL, HGSIL, and all others. We also included a dichotomous variable to indicate in which clinical practice the subject was seen.

Dependent Variables

The primary outcome for the study was timeliness of diagnostic resolution of the abnormal Pap test. Diagnostic resolution was defined as a definitive tissue diagnosis (biopsy with pathology), or a clinical evaluation determining that no further

Table 2. Subject Characteristics Before and After the Pap Tracking Intervention

	Pre-Intervention *N*=137 *n* (%)	Post-Intervention *N*=69 *n* (%)	*P*-value
Age			0.58
18–21	11 (8.0)	6 (8.7)	
22–26	33 (24)	22 (32)	
27–35	44 (32)	17 (25)	
36+	49 (36)	24(35)	
Race			0.85
White	57 (42)	30 (43)	
Black	53 (39)	24 (35)	
Other	27 (20)	15 (22)	
Language			0.66
English	122 (89)	60 (87)	
Non-English	15 (11)	9 (13)	
Insurance			0.16
Private	55 (40)	28 (41)	
Public	12 (8.8)	12 (17)	
Uninsured	70 (51)	29 (42)	
Abnormality			0.33
ASCUS/HPV+[a]	36 (26)	24 (35)	
AGC/AGUS[b]	5 (3.7)	4 (5.8)	
LGSIL[c]	85 (62)	39 (57)	
HGSIL[d]	6 (4.4)	2 (2.9)	
ASC-H[e]	5 (3.7)	0 (0)	

[a] *ASCUS = atypical squamous cells of undetermined significance; HPV = human papillomavirus*
[b] *AGC = atypical glandular cells; AGUS = atypical glandular cells of undetermined significance*
[c] *LGSIL = low-grade squamous intraepithelial lesion*
[d] *HGSIL = high-grade squamous intraepithelial lesion*
[e] *ASC-H = atypical squamous cells: cannot exclude high-grade squamous intraepithelial lesion*

evaluation was necessary[12]. We evaluated the outcome both as a dichotomous and continuous variable. For the dichotomous variable, we categorized subjects as to whether they had received diagnostic resolution by 365 days. For the continuous variable, we defined follow-up as the number of days to diagnostic resolution, top coding those who did not resolve to 366 days.

Data Analysis

Our primary research question was whether there was a difference between the pre-intervention and post-intervention groups. We calculated and tested for the differences in median time-to-resolution using the one-way Wilcoxon test, differences in mean time-to-resolution using the t-test and differences in percentages using the chi-squared test. We used the Cox-proportional hazard method for univariate and multivariate hazard analyses of likelihood of resolving within 365 days. From the model we then calculated an adjusted mean time-to resolution and confidence intervals. Those predictors that were significant at the $p<0.10$ level in the bivariate analyses were included in the final multivariate models. All analyses were conducted using SAS v9.1 (SAS Institute, Cary NC).

RESULTS

Subject Demographics

Table 2 shows demographic and Pap abnormality characteristics for subjects in the pre-intervention period (n=137) and post-intervention period (n=69); 2.2% of the total 9164 Pap tests performed during the study time periods were abnormal. There were no statistically significant differences between the pre- and post-intervention groups in age, race/ethnicity, primary language, insurance status, or type of Pap abnormality (all $p>0.10$). Overall, subject characteristics reflected the low income and minority populations cared for at our institution, with 60% of subjects publicly insured or uninsured, and more than 50% from a racial or ethnic minority group.

Outcomes

Table 3 shows the pre- and post-intervention results for the two primary measures of diagnostic resolution. The bivariate comparisons did not show a statistically significant difference between the pre- and post-intervention period, although the direction of difference favored our hypothesis. In bivariate hazard analyses we found that the severity of cytologic abnormality and practice location were associated at the $p<0.10$ level and therefore included in the multivariable analysis. Age, race/ethnicity, language, and insurance status were not associated with the outcome. After multivariate adjustment for type of Pap abnormality and practice location, subjects in the post-intervention period were significantly more likely to ever

Table 3. Outcomes of Abnormal Pap Tests Before and After Pap Tracking Intervention

	Pre-Intervention *N*=137	Post-Intervention *N*=69	*P*-value
Achieved resolution *n* (%)	127 (93%)	67 (97%)	0.20
MEDIAN days to resolution Median (IQR[a])	72 (47–112)	58 (36–102)	0.04
Unadjusted MEAN days to resolution Mean (CI)	108 (92–125)	86 (68–105)	0.11
Adjusted[b] *MEAN days to resolution Mean (CI)*	108 (105–112)	86(81–91)	0.0002
Adjusted odds ratio[b] *of EVER achieving resolution OR (95% CI)*	ref (1.00)	15.4 (3.7–62)	0.0002
Adjusted odds ratio[b] *of resolving in a SHORTER period of time OR (95% CI)*	ref (1.00)	1.40 (1.03–1.9)	0.03
Colposcopy result			
Non-neoplastic	60 (48%)	38 (57%)	0.63
CIN 1[c]	47 (38%)	19 (28%)	
CIN 2	11 (9%)	6 (9%)	
CIN 3	6 (5%)	3 (5%)	
Invasive cervical cancer	0 (0%)	1 (1%)	
Other	1 (<1%)	0 (0%)	

[a] *IQR = interquartile range*
[b] *Adjusted for type of Pap abnormality and practice location*
[c] *CIN = cervical intraepithelial neoplasia*

achieve diagnostic resolution (OR, 15.4; CI, 3.7–62) and more likely to achieve diagnostic resolution in a shorter period of time (HR, 1.40; CI, 1.03–1.9), relative to subjects in the pre-intervention period. The adjusted mean time to diagnostic resolution decreased from 108 days (CI 105–112 days) to 86 days (CI 81 to 91 days). Colposcopy results were similar between the two groups, (p=0.63) with 14% of women with CIN 2 or more severe lesions.

DISCUSSION

We developed and evaluated an EMR based tracking system for abnormal Pap tests, in order to assist providers in ensuring all abnormalities reached diagnostic resolution. We found that this tracking system significantly improved follow-up of abnormal Pap tests. Although many practices have developed electronic or paper tracking methods requiring manual entry, we report here on a novel method of incorporating the tracking within an electronic medical record that when initially developed had no intrinsic design features allowing this to happen.

Multiple studies have documented the problem of inadequate follow-up of abnormal Pap tests[3–11]. Leyden and colleagues looked at women diagnosed with cervical cancer from January 1995 to December 2000 in seven comprehensive health plans, and found that 13% of all cancers were attributed to inadequate follow-up of an abnormal Pap test[4]. Other studies have documented that 30–49% of women had either no follow up or delays beyond 3 to 7 months in abnormal Pap test follow-up in minority, uninsured or Medicaid-insured, or low income populations[3–11]. Systems specifically in safety net institutions that care for these communities are critical to improve follow-up rates and so improve effectiveness of cervical cancer screening. Our results are focused on resolution of abnormal Pap tests, and not in actual prevention of cervical cancer. Larger studies would be needed to demonstrate that such systems directly result in fewer women progressing to invasive cancer.

It is likely a combination of both systems and patient barriers that impede adequate abnormal Pap test follow-up. Patient barriers include difficulty in keeping follow-up appointments, limited understanding of the significance of the abnormality, and other life-issues taking priority. Systems barriers include failure of the provider to be aware of an abnormal result, and limited capability to systematically track patients who do not keep follow-up appointments. Our program addressed the systems barriers by giving providers tools to allow them to more easily track subjects after an abnormal Pap test. Our higher baseline follow up rates may already reflect some of the benefits of an EMR system; however, delays persisted without a tracking system.

Most of the EMRs used in outpatient medicine were developed with a focus on billing and require significant information technology development in data collection, synthesis and distribution to develop a tracking function. Our system required synthesis from multiple data sources, including scheduling, ordering, registration, and pathology. Data collection challenges included text-only fields in pathology reports; reports with standard result syntax or field based reports would avoid these pitfalls. Distribution of the reports was not automated, requiring that personnel adjust the programming and run the reports each month. Even during the intervention, changes to one part of the system resulted in difficulties and changes needed to access other parts of the system. Therefore, the system required some finite but constant resources to maintain, generate, and distribute the tracking reports. Despite these challenges, the system is in active use, and providers have reported satisfaction with the systems' ability to catch those cases that fall between the cracks. Most providers receive a monthly list of fewer than five patients. Of note, our work was supported through the medical center's risk management department, given the quality improvement benefits of the system. Given that many health care systems have adopted self insurance for malpractice, risk management funds might be a source to support other primary care initiatives for tracking and case management which improve quality of care and reduces risk[13].

One component of a medical home[14] is the ability to conduct population-based management of care, including tracking of abnormal screening test results. Our system serves as an example of the successful development and utilization of such a tracking system toward management of an entire practice. Our tracking system did not provide any additional assistance or personnel to the provider after they are given the monthly tracking report. Due to the relatively small number of abnormal Pap tests, providers were able to utilize existing staff personnel and their own efforts to ensure follow-up was achieved. This might be more of an issue if this tracking model was applied to a larger-volume abnormality, such as abnormal mammograms, cholesterol results, or glycosylated hemoglobin results in patients with diabetes. Additional resources, in terms of case management[15] or patient navigation[16], along with electronic tracking systems, have been employed to provide the additional follow through necessary when more frequent rates of an abnormality are expected, including phone or mail contact with patients, and rescheduling of missed appointments.

Acknowledgements: *Financial support for this study was provided by a grant from the Risk Management Fund at Boston Medical Center. This study was supported by Boston University School of Medicine, Boston, MA. Preliminary data were presented at the 2009 Annual Meeting of the Society of General Internal Medicine in Miami as a poster presentation.*

Conflict of Interest: *None disclosed.*

Corresponding Author: *Elizabeth A. Dupuis, MD; 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: Elizabeth.Dupuis@bmc.org).*

REFERENCES

1. The American Cancer Society. What Are the Key Statistics About Cervical Cancer? http://www.cancer.org/docroot/CRI/content/CRI_2_4_1X_What_are_the_key_statistics_for_cervical_cancer_8.asp. Accessed January 28, 2010.
2. **Schiffman MH, Brinton LA, Devesa SS.** Cervical cancer. In: **Schottenfeld D, Fraumeni JF, eds.** Cancer Epidemiology and Prevention. 2nd ed. New York: Oxford University Press; 1996:1090–116.
3. **Peterson NB, Han J, Freund KM.** Inadequate follow-up for abnormal Pap smears in an urban population. J Natl Med Assoc. 2003;95(9):825–32.

4. **Leyden WA, Manos MM, Geiger AM, et al.** Cervical cancer in women with comprehensive health care access: attributable factors in the screening process. J Natl Cancer Inst. 2005;97(9):675–83.
5. **Marcus AC, Crane LA, Kaplan CP, et al.** Improving adherence to screening follow-up among women with abnormal Pap smears: results from a large clinic-based trial of three intervention strategies. Med Care. 1992;30(3):216–30.
6. **Cartwright PS, Reed G.** No-show behavior in a county hospital clinic. Am J Gynecol Health. 1990;6:15–21.
7. **Laedtke TW, Dignan M.** Compliance with therapy for cervical dysplasia among women of low socioeconomic status. South Med J. 1992;85(1):5–8.
8. **Lane DS.** Compliance with referrals from a cancer-screening project. J Fam Pract. 1983;17(5):811–7.
9. **Carey P, Gjerdingen DK.** Follow-up of abnormal Papanicolaou smears among women of different races. J Fam Pract. 1993;37(6):583–7.
10. **Nathoo V.** Investigation of non-responders at a cervical cancer screening clinic in Manchester. Br Med J (Clin Res Ed). 1988;296(6628):1041–2.
11. **Paskett ED, White E, Carter WB, Chu J.** Improving follow-up after an abnormal Pap smear: a randomized controlled trial. Prev Med. 1990;19 (6):630–41.
12. **Freund KM, Battaglia TA, Calhoun E, et al.** National Cancer Institute Patient Navigation Research Program: methods, protocol, and measures. Cancer. 2008;113(12):3391–9.
13. **Townsend RW.** Formation and operation of a captive insurance company for malpractice coverage. Coll Rev. 1988;5:47–61.
14. National Committee for Quality Assurance (NCQA). Standards and Guidelines for Physician Practice Connections - Patient-Centered Medical Home (PPC-PCMH): 2008.
15. **Norris SL, Nichols PJ, Caspersen CJ, et al.** The effectiveness of disease and case management for people with diabetes. A systematic review. Am J Prev Med. 2002;22(4 Suppl):15–38.
16. **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.

The Effectiveness of Tax Policy Interventions for Reducing Excessive Alcohol Consumption and Related Harms

Randy W. Elder, PhD, Briana Lawrence, MPH, Aneeqah Ferguson, MPA, Timothy S. Naimi, MD, MPH, Robert D. Brewer, MD, PhD, Sajal K. Chattopadhyay, PhD, Traci L. Toomey, PhD, Jonathan E. Fielding, MD, MPH, MBA, the Task Force on Community Preventive Services

Abstract: A systematic review of the literature to assess the effectiveness of alcohol tax policy interventions for reducing excessive alcohol consumption and related harms was conducted for the *Guide to Community Preventive Services* (*Community Guide*). Seventy-two papers or technical reports, which were published prior to July 2005, met specified quality criteria, and included evaluation outcomes relevant to public health (e.g., binge drinking, alcohol-related crash fatalities), were included in the final review. Nearly all studies, including those with different study designs, found that there was an inverse relationship between the tax or price of alcohol and indices of excessive drinking or alcohol-related health outcomes. Among studies restricted to underage populations, most found that increased taxes were also significantly associated with reduced consumption and alcohol-related harms. According to *Community Guide* rules of evidence, these results constitute strong evidence that raising alcohol excise taxes is an effective strategy for reducing excessive alcohol consumption and related harms. The impact of a potential tax increase is expected to be proportional to its magnitude and to be modified by such factors as disposable income and the demand elasticity for alcohol among various population groups.

(Am J Prev Med 2010;38(2):217–229) Published by Elsevier Inc. on behalf of American Journal of Preventive Medicine

Introduction

Excessive alcohol consumption is the third-leading actual cause of death in the U.S.,[1] and each year it accounts for approximately 79,000 deaths and 2.3 million years of potential life lost (about 29 years of life lost per death; apps.nccd.cdc.gov/ardi/Homepage.aspx). Excessive alcohol consumption contributes to a variety of health and social problems, including unintentional injuries (e.g., injuries due to motor vehicle crashes); suicide; homicide; liver cirrhosis; gastrointestinal cancers; vandalism; and lost productivity.[2] Alcohol consumption by underage drinkers also contributes to the three leading causes of death among adolescents (unintentional injuries, suicide, and homicide),[3] and any underage drinking is considered excessive.

One of the fundamental laws of economics is that quantity demanded of a product is inversely related to its price (Law of Demand).[4] Based on economic theory, therefore, increasing the price of alcohol would be expected to lower alcohol consumption. Alcohol taxes are promulgated primarily by federal and state governments, but can be instituted at the local or county level. Currently in the U.S., alcohol taxes are beverage-specific (i.e., they differ for beer, wine, and distilled spirits) and are usually "nominal" taxes, meaning they are based on a set rate per unit volume and are not adjusted for inflation (i.e., they generally remain stable as the cost of living increases). At the state and federal levels, inflation-adjusted alcohol taxes have declined considerably since the 1950s.[5] Concordant with this decrease in the real value of these taxes from substantially higher levels, the inflation-adjusted price of alcohol decreased dramatically,[6] reflecting the

From the National Center for Health Marketing (Elder, Lawrence, Ferguson, Chattopadhyay), and National Center for Chronic Disease Prevention and Health Promotion (Naimi, Brewer), CDC, Atlanta, Georgia; University of Minnesota School of Public Health (Toomey), Minneapolis, Minnesota; and Los Angeles County Department of Public Health (Fielding), Los Angeles, California

Address correspondence and reprint requests to: Randy W. Elder, PhD, Guide to Community Preventive Services, 1600 Clifton Road, Mailstop E-69, Atlanta GA 30333. E-mail: rfe3@cdc.gov.

0749-3797/00/$17.00
doi: 10.1016/j.amepre.2009.11.005

fact that changes in taxes are efficiently passed on through changes in prices.[7] The goal of this systematic review is to assess the relationship between alcohol taxes or prices and public health outcomes related to excessive alcohol consumption to better inform decision makers about the potential utility of using tax policy as a means of improving those outcomes.

Healthy People 2010 Goals and Objectives

The intervention reviewed here is relevant to several objectives specified in *Healthy People 2010*, the disease prevention and health promotion agenda for the U.S. (Table 1).[8] The objectives most directly relevant to this review are those that aim to reduce excessive alcohol consumption (26-11); reduce average annual alcohol consumption (26-12); and reduce key adverse consequences of excessive alcohol consumption (26-1, 26-2, and 26-5 through 26-8). In addition to these specific objectives, *Healthy People 2010* notes that excessive alcohol consumption is also related to several other public health priorities such as cancer, educational achievement, injuries, risky sexual activity, and mental health; thus, a reduction in excessive alcohol consumption should help to meet some of the national goals in these areas as well.

Table 1. Selected *Healthy People 2010*[8] objectives related to excessive alcohol consumption

Adverse consequences of substance use and abuse
26-1 Reduce alcohol-related motor-vehicle fatalities[a]
26-2 Reduce cirrhosis deaths
26-5 Reduce alcohol-related hospital emergency department visits
26-6 Reduce the proportion of adolescents who ride with drinking drivers
26-7 Reduce intentional injuries resulting from alcohol-related violence[a]
26-8 Reduce cost of lost productivity due to alcohol use[a]
Substance use and abuse
26-10a Increase proportion of adolescents not using alcohol in past 30 days[a]
26-11 Reduce proportion of people[b] engaging in binge drinking
26-12 Reduce average annual alcohol consumption
26-13 Reduce proportion of adults who exceed guidelines for low-risk drinking

[a]Objective also relates to illicit drug use
[b]Aged ≥12 years

Recommendations from Other Advisory Groups

Several authors[9–12] have suggested that increasing alcohol prices by raising alcohol excise taxes is among the most effective means of reducing excessive drinking and alcohol-related harms. Increasing alcohol excise taxes has been specifically recommended as a public health intervention by the IOM, Partnership for Prevention, the WHO, and the expert panel convened for the Surgeon General's Workshop on Drunk Driving.[13–16] These recommendations are based on studies[14,17,18] showing that increased alcohol taxes are associated with decreased overall consumption, decreased youth consumption, decreased youth binge drinking, reduced alcohol-related motor-vehicle crashes, reduced mortality from liver cirrhosis, and reduced violence.

The Guide to Community Preventive Services

The current systematic review of the effects of alcohol taxes and prices on excessive alcohol consumption and related harms applies the stringent inclusion and assessment criteria of the *Guide to Community Preventive Services* (*Community Guide*).[19] It was conducted under the oversight of the independent, nonfederal Task Force on Community Preventive Services (Task Force), with the support of USDHHS in collaboration with public and private partners. The CDC provides staff support to the Task Force for development of the *Community Guide*.

To support efforts to address important public health priorities, such as reducing excessive alcohol consumption and its related harms, the Task Force makes recommendations for practice and policies based on the results of *Community Guide* reviews such as this one. These recommendations are based primarily on the effectiveness of an intervention in improving important outcomes as determined by the systematic literature review process. In making its recommendations, the Task Force balances information about effectiveness with information about other potential benefits and harms of the intervention itself. The Task Force also considers the applicability of the intervention to various settings and populations in determining the scope of the recommendation. Finally, the Task Force reviews economic analyses of effective interventions, where available. Economic information is provided to assist with decision making, but it generally does not affect Task Force recommendations. See the Task Force–authored paper in this issue for recommendations regarding the effects of alcohol taxes and prices on excessive alcohol consumption and related harms.[20]

Evidence Acquisition

Community Guide methods for conducting systematic reviews and linking evidence to effectiveness are described elsewhere[19] and on the *Community Guide* website (www.thecommunityguide.org/methods). In brief, for each *Community Guide* review topic, a systematic review development team representing diverse disciplines, backgrounds, and work settings conducts a review by (1) developing a conceptual approach to identify, organize, group, and select interventions for review; (2) developing a conceptual model depicting interrelationships among interventions, populations, and outcomes; (3) systematically searching for and retrieving evidence; (4) assessing and summarizing the quality and strength of the body of evidence of effectiveness; (5) translating evidence of effectiveness into recommendations; (6) summarizing data about applicability (i.e., the extent to which available effectiveness data might apply to diverse population segments and settings), economic impact, and barriers to implementation; and (7) identifying and summarizing research gaps.

Conceptual Model

The conceptual causal pathway by which increased alcohol taxes are expected to reduce excessive alcohol consumption and its related harms is depicted in Figure 1. The first step in this pathway posits that tax increases will be passed on to the consumer in the form of higher alcohol prices, as has been documented previously.[7] According to the Law of Demand,[4] an increased price would be expected to lead to a decrease in the quantity of alcoholic beverages demanded, resulting in decreases in excessive alcohol consumption and its harmful consequences. Details of the specific independent variables and outcome measures that reflect the concepts in this conceptual causal pathway are provided below.

One complicating factor in this conceptual model arises from the fact that different types of alcoholic beverages (e.g., beer, wine, and spirits) are taxed at different rates in the U.S. and several other countries. When tax increases affect one type of beverage only (designated as the "targeted" alcoholic beverage in Figure 1), one must consider the possibility of substitution effects, whereby alcoholic beverages that have not been affected by the tax increase may be consumed in greater quantities. To the extent that such substitution occurs, the overall rate of excessive drinking would not decrease as much as would otherwise be expected based on the decrease in quantity demanded for the beverage targeted by the tax increase. However, binge drinkers are known to prefer certain types of alcoholic beverages (e.g., most adult binge drinkers in the U.S. consume beer)[21] for reasons that may not be entirely related to price (e.g., availability, convenience, taste); thus, it is not clear whether and how large an effect beverage substitution would likely have on overall alcohol consumption, even when tax increases affect one beverage type only.

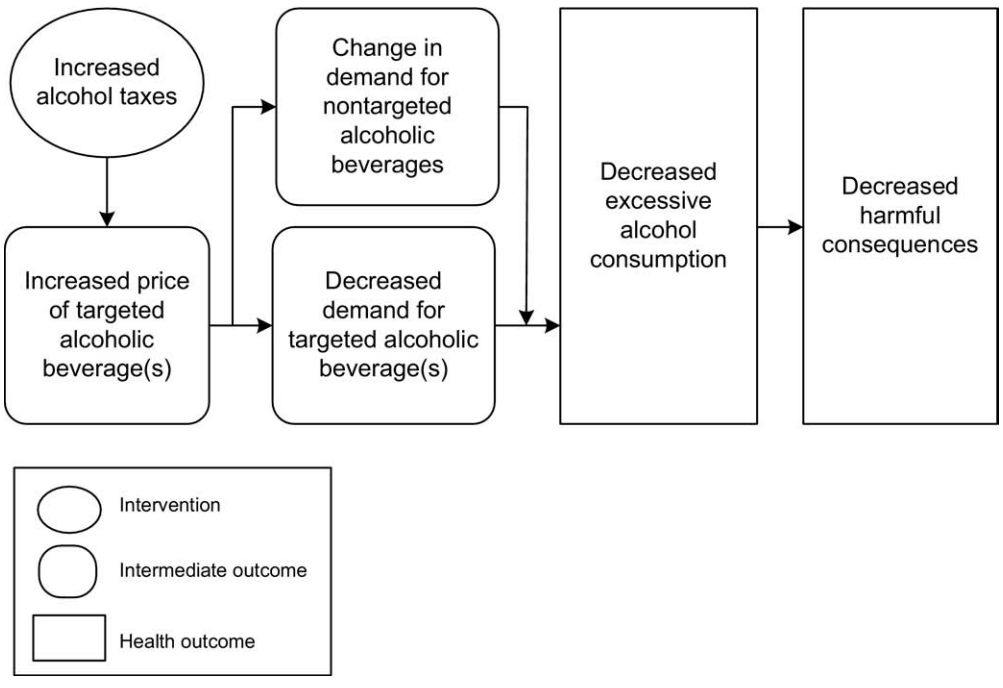


Figure 1. Conceptual model for the causal relationship between increased alcohol taxes and decreased excessive alcohol consumption and related harms (oval indicates intervention; rectangles with rounded corners indicate mediators or intermediate outcomes; and rectangles indicate outcomes directly related to improved health)

Review Inclusion Criteria

To be considered for inclusion in this review, candidate studies had to (1) meet minimum *Community Guide* standards for study design and quality[19]; (2) be published in an English-language journal, book chapter, or technical report; (3) be conducted in a high-income economy; and (4) evaluate independent variables and outcome measures of interest.

Independent variables of interest. In addition to the other criteria noted above, to be included in this review, a study had to evaluate either the effects of a

change in alcohol tax policy or the relationship between alcohol taxes or prices and outcomes of interest. Studies of the effects of alcoholic beverage prices were considered relevant to an evaluation of alcohol taxes because there is evidence that changes in alcohol taxes are passed on to the consumer in the form of higher or lower prices, with little or no lag time.[7] In fact, there is some evidence that tax increases may be magnified as they are passed on to the consumer. For example, when the federal excise tax on beer increased by $9 per barrel in 1991, it was estimated to have increased retail prices by $15 to $17.[7]

Outcome measures of interest. The outcome measures of interest in this review are direct measures or proxies relating to the two final boxes in Figure 1—that is, excessive alcohol consumption and the harmful consequences of such consumption. When excessive alcohol consumption is assessed directly, it is typically done through surveys assessing either the prevalence or frequency of binge drinking (four or more drinks per occasion for women, or five or more drinks per occasion for men); heavy drinking (more than seven drinks per week for women, or more than 14 drinks per week for men); or underage drinking (defined by state or national laws). Measures of societal levels of alcohol sales or consumption were also considered an acceptable proxy for excessive consumption for two primary reasons. First, there is an extremely strong relationship between per capita alcohol consumption and various measures of excessive drinking.[22,23] Furthermore, because people consuming greater quantities of alcohol may be more sensitive to price increases, reductions in societal levels of alcohol consumption subsequent to price increases may result in even larger declines in excessive consumption.[22]

In addition to studies directly or indirectly assessing excessive alcohol consumption, studies assessing health-related outcomes associated with excessive alcohol consumption (e.g., alcohol-related motor-vehicle crashes) were also included in this review. In some cases, a single paper reported multiple measures of a single general outcome (e.g., both single-vehicle nighttime crashes and total crashes reported as measures of alcohol-related crashes). In these instances, the measure that was most strongly associated with excessive alcohol consumption based on estimated alcohol-attributable fractions was chosen as the primary result reported for that outcome.

Search for Evidence

Conducting a thorough search for studies of the effects of alcohol taxes or alcohol prices is challenging because the effects of alcohol taxes or prices are often studied in conjunction with many other variables. As a result, a search that targets "tax" or "price" may fail to identify many relevant studies. To address this issue, a search was conducted for relevant studies as part of a broad database search for terms related to several alcohol policy interventions of interest to the current review group, covering the period from database inception through July 2005. Using MeSH terms and text words, the following databases were searched: MEDLINE, EMBASE, PsycINFO, the ETOH database of the National Institute on Alcohol Abuse and Alcoholism, Web of Science, Sociological Abstracts, and EconLit. Search strategies are available at www.thecommnityguide.org/alcohol/supportingmaterials/SSincreasingtaxes.html. The reference lists of prior literature reviews, as well as reference lists from studies included in this review, were used to identify additional relevant articles. The search produced 5320 potentially relevant papers, of which 78 met the inclusion criteria.

Data Extraction and Quality Assessment

For each candidate study, study characteristics and results were recorded, and the quality of study execution was assessed. The degree to which a study's basic design protected against threats to internal validity was rated using a three-level classification system ranging from least suitable (for designs with a cross-sectional analysis or a single observation before and after an intervention) to greatest suitability (for designs with concurrent comparison conditions).[19] Ratings of the quality of each study's execution provided further information on their utility for the purposes of the review. Quality of study execution was assessed using a standard 9-point scale, reflecting the total number of identified limitations to internal or external validity (viz. study population and intervention descriptions, sampling, exposure and outcome measurement, data analysis, interpretation of results, and other biases). Studies with zero or one limitation were categorized as having good execution, those with two to four limitations had fair execution, and those with five or more limitations were categorized as having limited execution.[19] Studies with limited execution were excluded from further analysis.

Effect Measurement and Synthesis of Results

The most common method for studying the effects of alcohol taxes on alcohol-related outcomes is to assess how they (or the prices they influence) relate to those outcomes over time, while controlling for potential confounding factors. For most of the studies in this review, the reported results were either directly reported as elasticities or were transformed into elasticities. These were then directly compared with elasticities calculated from other studies. An elasticity represents the percentage change in a dependent variable associated with a 1% increase in an independent variable (e.g., price or tax rate). For example, a price elasticity of −0.5 means that a 10% increase in price would be expected to result in a 5% decrease in the outcome of interest. Tax elasticities have a similar interpretation, but cannot be directly compared with price elasticities because taxes represent only a fraction of the total purchase price (resulting in smaller values for tax elasticities). In most cases for which

elasticities were not reported in the original studies, only the direction and significance of the reported effects could be evaluated in this review.

Because elasticities are measures of relative change, they provide a common metric for comparing and aggregating related, but not identical, outcomes (e.g., different measures of alcohol consumption; different types of motor-vehicle crashes). In general, measures of alcohol consumption fell into two broad categories: those that evaluate indices of consumption at the societal level (e.g., total alcohol sales) and those that evaluate consumption at the individual level (e.g., self-reported binge drinking). Measures of alcohol-related harms were grouped into broad categories of related outcomes, such as motor-vehicle crashes, liver cirrhosis, violence, alcohol dependence, and all-cause mortality.

For most of the outcomes of interest in this review, results were synthesized descriptively, without the use of any summary effect measures, due to a substantial amount of variation in the specific outcomes assessed and in the units used to measure the effects of changes in taxes or prices. The only outcome for which both enough studies and sufficiently similar results were found to allow a quantitative synthesis of the results was societal-level alcohol consumption. Data from these studies were summarized graphically and by using descriptive statistics, specifically medians and interquartile intervals. These results were also stratified on several variables considered by the review team to be potentially important effect modifiers (e.g., study design), allowing for an assessment of the robustness and generalizability of the results. This approach to synthesis was primarily chosen for the following two reasons. First, because many of the included studies had some overlap with respect to the locations and time periods covered in their analyses, their results were not completely independent. Second, many of these studies did not report results in a way that allowed for the calculation of CIs for their elasticities.

For studies that reported stratified results (e.g., separate price elasticities for beer, wine, and spirits), the median value across the relevant strata reported in that study was used for the calculation of summary statistics. This approach prevented studies that reported multiple outcomes from having undue influence on the summary statistics.

Evidence Synthesis

Description of Included Studies

A total of 78 papers[24–101] reported on studies that met the review inclusion criteria. Only some of the outcomes from one study[83] were included because not all of its analyses met quality of execution criteria. Five other studies[70,88–91] were excluded from the review because they failed to meet quality of execution criteria. Detailed descriptions of the included studies are available at www.thecommunityguide.org/alcohol/supportingmaterials/SETincreasingtaxes.html.

Most studies assessed total alcohol consumption at the societal level (i.e., per capita alcohol consumption). The design of these studies varied across countries. Most studies conducted outside the U.S. used interrupted time–series designs, because alcohol taxes in other countries tend to be set at the national level, and as such, it is generally not possible to do intra-country comparisons. In contrast, most of the U.S. studies used a panel study design, in which multiple states were assessed over time, allowing each to serve as a comparison for the others. These studies included both those that accounted for between-state differences using a fixed-effects approach (whereby stable between-state differences are controlled for by dummy coding) and those that used a random-effects approach (whereby between-state differences in variables other than tax or price are controlled for by including important predictors of alcohol consumption in the model). The remaining studies assessed measures related to excessive drinking (e.g., the prevalence of underage or binge drinking) or alcohol-related harms, the most common being outcomes related to motor-vehicle crashes.

Intervention Effectiveness

Alcohol price and overall consumption. Of the studies in the review, 50 assessed overall alcohol consumption; 38 (76%) of these reported price elasticities[25,27,33–38,40,43,45,47,48,52,53,57,63,65,67,71,73,74,77,78,80–82,84,92–95,97] (six of these studies came from one paper[80] that calculated elasticities for multiple countries). Almost all of these 38 studies (95%) reported negative price elasticities, indicating that higher prices were associated with lower consumption. These results were quite consistent across beverage type, with median elasticities ranging from −0.50 for beer to −0.79 for spirits (Figure 2). Similarly, interquartile intervals for beer, wine, and spirits were also consistent across beverage type, with the 25th percentile elasticity ranging from −0.91 to −1.03, and the 75th percentile ranging from −0.24 to −0.38. Results for studies of overall ethanol consumption across beverage types were somewhat more variable because of the presence of several outliers with very large elasticities; for this outcome, the 75th percentile was comparable to that for the other outcomes (−0.50), but the 25th percentile had a substantially larger absolute value (−2.00).

As indicated in Table 2, the price elasticities reported in the reviewed studies were also quite consistent when evaluated by study characteristics (i.e., design suitability, model type, time period, and location). Across all of the nine strata examined, median elasticities ranged from −0.51 to −0.90, the 25th percentile elasticities ranged

from −0.78 to −1.10, and the 75th percentile elasticities ranged from −0.32 to −0.50. The most notable differences in elasticities across strata were among panel studies that used fixed- versus random-effects regression models. In general, fixed-effects models tended to produce elasticities of slightly smaller magnitude than did random-effects models. This might be expected because the elasticities from fixed-effects models do not account for between-state differences in taxes that are stable over time (although these models have several other desirable qualities).

Of the 50 studies that assessed overall alcohol consumption, 12 studies[29,31,32,39,41,49,54,75,76,83,98,99] assessed the relationship between price and overall consumption, but these studies did not provide price elasticities or sufficient information to calculate them. Many of these studies reported the results of multiple analyses that produced separate results for different subpopulations, beverage types, or analytic models with different parameters. In eight of these studies,[29,31,32,39,41,54,76,83] all of the reported results indicated that higher prices were associated with lower alcohol consumption; in seven,[29,31,32,39,41,54,83] results were significant across all analyses, and one[76] had results of mixed significance across analyses. The other four studies[49,75,98,99] had mixed results across beverage types or analytic models, with some results in the expected direction and some in the opposite direction.

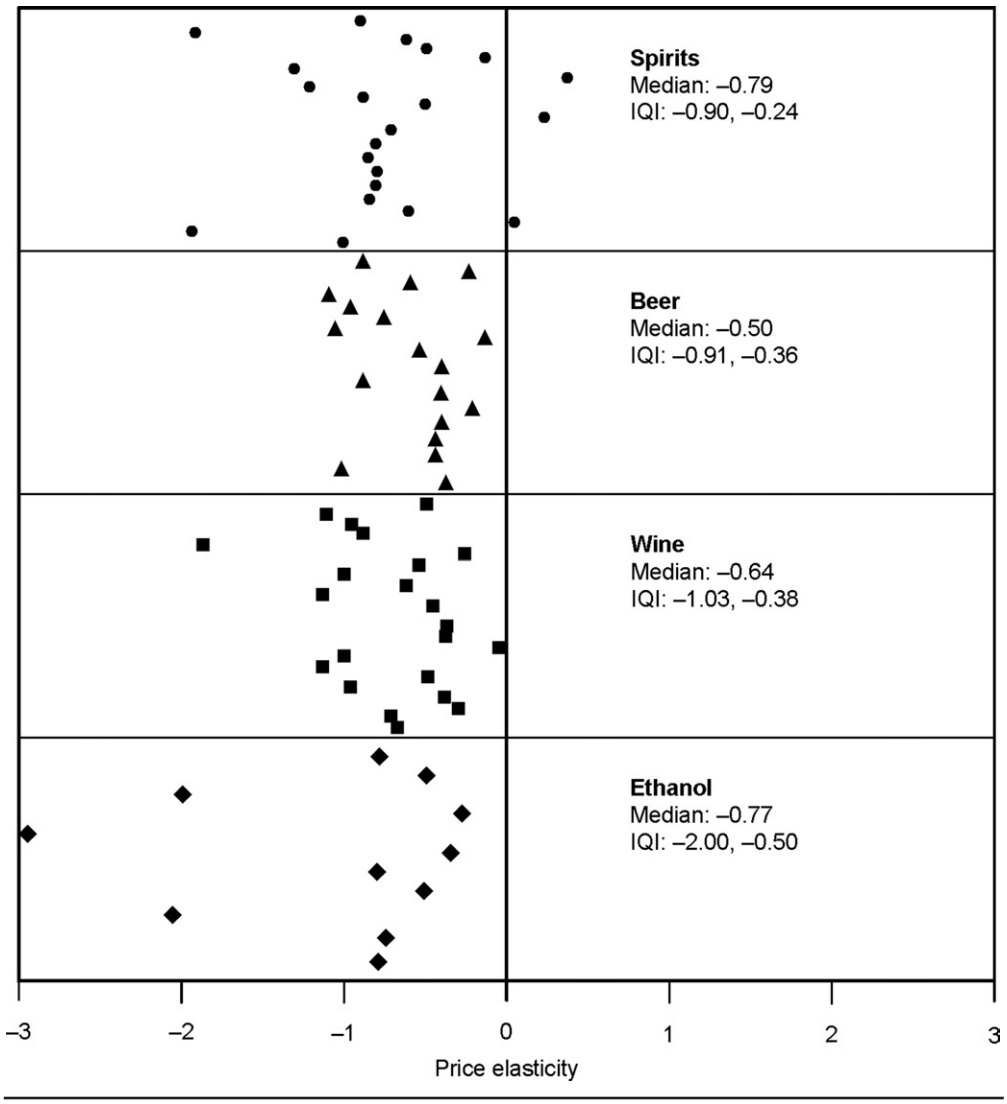


Figure 2. Scatterplot showing the association between alcohol price elasticities and excess consumption as measured by societal alcohol consumption. Each data point represents a single study's elasticity estimate for the given beverage type.
IQI, interquartile interval

Alcohol price or taxes and individual consumption patterns. Sixteen studies[24,46,53–56,58–62,64,68,72,96,102] in the review used survey data to evaluate the effects of alcohol prices or taxes on individual alcohol consumption patterns. Most of these studies assessed the prevalence of alcohol consumption among youth aged <25 years, primarily underage youth. Respondent groups included high school students, college students, young people in the general population, and adults in the general population. All but two of these studies[54,59] were conducted in the U.S.

Of the nine studies[24,46,56,58,60–62,64,68] that assessed the relationship between alcohol price or taxes and drinking prevalence among young people, six[46,56,58,60,61,68] consistently indicated that higher prices or taxes were associated with a lower prevalence of youth drinking (four with one or more significant findings). Three of these studies reported price elasticities: −0.29 for drinking among high school students;[46] −0.53 for heavy drinking among

Table 2. Medians and interquartile intervals for price elasticity of alcohol consumption, stratified by study characteristics

Characteristic (no. of studies)	Median elasticity	Interquartile interval
Design suitability		
Greatest suitability (16)	−0.76	−1.06 to −0.50
Moderate suitability (16)	−0.51	−0.85 to −0.39
Least suitable (6)	−0.68	−0.94 to −0.32
Model type		
Random effects (7)	−0.90	−1.10 to −0.50
Fixed effects (8)	−0.69	−0.78 to −0.40
Time period[a]		
Before 1963 (19)	−0.61	−0.90 to −0.38
1963 or later (19)	−0.76	−0.89 to −0.44
Location		
U.S. (21)	−0.63	−0.90 to −0.44
Non-U.S. (17)	−0.68	−0.88 to −0.37

[a]First data point in time–series

those aged 16–21 years[58]; and −0.95 and −3.54, respectively, for binge drinking among men and women aged 18–21 years.[61] The three remaining studies[24,62,64] reported mixed results across different analyses, with the majority of their effect estimates indicating an inverse relationship between tax or price and drinking observed in the studies above.

The nine studies that assessed the relationship between price or taxes and alcohol consumption patterns in adults or in the general population also generally found that increasing the prices or taxes on alcoholic beverages was associated with a lower prevalence of excessive alcohol consumption and related harms. Two of these studies assessed the relationship between alcohol price and the prevalence of binge drinking using data from the National Longitudinal Survey of Youth, which followed a group of people aged 14–22 years in 1979.[55,68] In a cohort of those aged 25–26 years from this survey, higher prices were associated with significant decreases in both overall alcohol consumption and frequent binge drinking (more than four episodes per month).[68] However, in a subsequent study of a cohort of those aged 29–33 years, higher prices were not significantly associated with the overall prevalence of binge drinking, and the direction of effects varied across beverage types.[55] Other studies based on surveys of the general adult population found that higher alcohol prices were associated with a lower overall prevalence of current drinking[72] and binge drinking,[53,72,102] and with a lower frequency of binge drinking.[53,72,96,102]

Three studies reported elasticities for the relationship between price and binge drinking; these ranged from −0.29 to −1.29, levels that are comparable to those for overall societal-level consumption.[53,61,96] Two additional studies evaluated a tax change in Switzerland that resulted in a 30% to 50% decrease in the price of imported spirits.[54,59] These studies found that the change was associated with a small (2.3%) increase in the prevalence of any drinking, and larger increases in measures of excessive alcohol consumption, specifically binge drinking (3.4%) and heavy drinking (9.3%). It is also noteworthy that the most marked increases in spirits consumption occurred among young men.

In summary, most studies that were included in this review found that higher taxes or prices were associated with reductions in alcohol consumption in general and excessive alcohol consumption in particular. Although these effects were not restricted to a particular demographic group, there is some evidence that they may be more pronounced among groups with a higher prevalence of excessive alcohol consumption (e.g., young men).

Alcohol price or taxes and alcohol-related harms. Twenty-two studies in the review evaluated the effects of changes in alcohol price[28,44,51,61,72,83,93,100] or taxes[24–26,29–31,66,69,85–87,98,101,103] on various alcohol-related harms. The most common outcomes evaluated were motor-vehicle crashes (including crash fatalities), various measures of violence, and liver cirrhosis. The studies were primarily conducted in the U.S., using state-level data.

Motor-vehicle crashes and alcohol-impaired driving. Eleven studies evaluated the effects of alcohol price[44,72,93,100] or taxes[24,26,29,30,86,98,103] on motor-vehicle crashes (Table 3). These studies found that the relationship between alcohol prices or taxes and injuries and deaths due to motor-vehicle crashes was generally significant and of a comparable magnitude to the relationship between these variables and alcohol consumption. The numeric values of the reported elasticities are substantially higher for studies that assessed the effects of alcohol prices than for those that assessed changes in alcohol taxes. This reflects the fact that taxes represent a relatively small proportion of the total purchase price of alcoholic beverages, so a larger proportional increase in taxes is necessary to achieve the same effect on the final purchase price of alcoholic beverages as a smaller proportional increase in the price itself. The reported elasticities were also generally higher for studies that assessed outcomes more directly attributable to alcohol consumption (e.g., alcohol-related crashes) than to those for which the relationship to alcohol consumption was less direct (e.g., all crash fatalities).

Three studies evaluated the relationship between alcohol prices[44,61] or taxes[66] and self-reported alcohol-impaired driving. These studies consistently found that alcohol-impaired driving was inversely related to the price of alcoholic beverages. The estimated price elasticities were similar for samples of Canadian[44] and U.S.[61] adults (range of −0.50 to −0.81; all $p<0.05$). The U.S. study stratified their sample by age in addition to gender, and reported price elasticities of −1.26 to −2.11 (both with $p<0.05$) for men and women aged 18–21 years, respectively.[61] The estimated tax elasticities from the remaining study were substantially larger for women than men (−0.29 vs −0.06), but neither estimate was significant.[66]

Table 3. Results of studies evaluating the relationship between alcohol prices or taxes and motor-vehicle crashes

Study	Independent variable	Dependent variable	Elasticity (*p*-value)
Price elasticity studies			
Cook (1981)[93]	Ethanol price[a]	Fatalities	−0.70 (NR)
Adrian (2001)[44]	Ethanol price[a]	Alcohol-related crashes	−1.20 (<0.05)
Sloan (1994)[72]	Ethanol price[a]	Fatalities	<0 (>0.05)
Whetten-Goldstein (2000)[100]	Ethanol price[a]	Alcohol-related fatalities	<0 (>0.05)
Tax elasticity studies			
Chaloupka (1993)[26]	Beer tax	Alcohol-related fatalities, all ages	−0.097 (<0.05)
	Beer tax	Alcohol-related fatalities, youth aged 18–20 years	−0.156 (<0.05)
Evans (1991)[86]	Beer tax	Single-vehicle nighttime fatalities	−0.12 (<0.05)
Ruhm (1996)[30]	Beer tax	Nighttime fatalities, youth aged 15–24 years (by age)	−0.18 (<0.05)
Saffer (1987)[42]	Beer tax	Fatalities, youth aged 15–24 years (by age)	−0.18 to –0.27 (all <0.05)
Ruhm (1995)[29]	Beer tax	Fatalities	<0 (<0.05)
Mast (1999)[98]	Beer tax	Fatalities	<0 (>0.05)
Dee (1999)[24]	Beer tax	Nighttime fatalities, youth aged 18–20 years	>0 (>0.05)

[a]Average price per ounce of ethanol across beer, wine, and spirits

Non–motor-vehicle mortality outcomes. Six studies evaluated the effects of alcohol price[25,28,72,83,93] or taxes[31] on nontraffic deaths. Despite substantial variability in their individual effect estimates, all six studies found that higher alcohol prices were associated with decreased mortality.

Five studies evaluated the relationship between alcohol prices and deaths from liver cirrhosis.[25,28,72,83,93] The two studies that reported results as elasticities produced substantially different elasticity estimates for this outcome, −0.90 ($p<0.05$)[93] and −0.01 ($p>0.05$).[28] Results of another study indicated that a $1 increase in the spirits tax would lead to a 5.4% decrease in cirrhosis ($p<0.05$).[25] Another found a nonsignificant effect in the expected direction.[72] The final study found a strong correlation of −0.87 between alcohol prices and cirrhosis deaths.[83] Although all of these studies indicate a consistent relationship between higher prices and lower cirrhosis mortality, there are substantial differences in the estimated strength of this relationship, which may be due to methodologic differences among studies.

One of the studies that evaluated cirrhosis mortality also assessed the relationship between alcohol price and several other causes of death.[72] The researchers found that there was a significant ($p<0.05$) inverse relationship between the price of alcoholic beverages and deaths from alcohol-related cancers (e.g., breast cancer) and suicide, and a nonsignificant ($p>0.05$) relationship between alcohol prices and deaths from homicides, falls, fires/ burns, and other injuries. Although these findings are surprising given the stronger relationship between alcohol consumption and intentional and unintentional injuries, the findings were robust across several regression models.

One study assessed all-cause mortality using a two-stage process.[31] In the first stage, the authors assessed the relationship between alcohol taxes and sales, and found that a one-cent increase in taxes per ounce of ethanol (a tax increase of approximately 10%) would be expected to result in a 2.1% decrease in sales. In the second stage, they found that a 1% decrease in alcohol sales was associated with a 0.23% decrease in all-cause mortality rates ($p<0.05$).

Violence outcomes. Three additional studies found that higher alcohol taxes are associated with decreased violence.[69,85,101] When the differences among tax and price elasticities are taken into account, the strength of the relationships reported in these studies were comparable to those found for alcohol consumption outcomes. The first

study estimated that beer tax elasticities on violent crime rates in the U.S. were −0.03 ($p>0.05$) for homicide; −0.03 ($p>0.05$) for assault; −0.13 ($p<0.05$) for rape; and −0.09 ($p<0.05$) for robbery.[101] The other two studies assessed the relationship between beer taxes and violence toward children, with different methods using overlapping samples. In the first analysis,[69] tax elasticities were −0.12 ($p<0.05$) for any violence toward children and −0.16 ($p<0.10$) for severe violence toward children. The subsequent analysis found that these results appeared to be due to an influence of taxes on violence by women but not by men.[85]

Other outcomes. Two studies evaluated the association between alcohol prices and two other health-related outcomes: alcohol dependence and sexually transmitted diseases. The first estimated an alcohol price elasticity for alcohol dependence of −1.49 ($p<0.05$).[51] The second used multiple methods of evaluating the effect of tax changes on sexually transmitted diseases, and found robust effects on rates of both gonorrhea and syphilis.[87]

Applicability

The Law of Demand[4] states that the inverse relationship between the price of a commodity and the quantity demanded is almost universal, and that only the strength of this relationship will vary across commodities or population groups. Consistent with these expectations, estimates of price elasticity for societal levels of alcohol consumption were robust across the various high-income economies in North America, Europe, and the Western Pacific Region evaluated in the studies in this review. Although results for harms related to excessive consumption came primarily from the U.S. and Canada, these findings are likely to be broadly applicable across high-income countries.

One important factor hypothesized to affect the strength of price elasticities for alcohol across different population groups is disposable income. Specifically, groups with less disposable income, such as underage drinkers, may be expected to be more sensitive to changes in alcohol prices than those with more disposable income.[104] Unfortunately, based on the studies in this review, it was not possible to determine whether alcohol price elasticities differ significantly on the basis of age or income. Furthermore, although the reviewed studies provided evidence that changes in alcohol prices affect excessive consumption (e.g., the prevalence and frequency of binge drinking), the available data were not adequate to assess potential differences in price elasticities based on drinking pattern (i.e., between excessive and nonexcessive drinkers).

Economic Efficiency

Our systematic economic review identified two studies that estimated the cost effectiveness of alcohol tax intervention based on modeling.[10,105] The first study[105] assessed the costs and outcomes of 84 injury prevention interventions for the U.S. and found that an alcohol tax of 20% of the pretax retail price offered net cost savings (i.e., the savings outweigh the costs) even after taking into account the adverse economic impact of reduced alcohol sales. The second study[10] analyzed the comparative cost effectiveness of alternative policies to reduce the burden of hazardous alcohol use for 12 WHO subregions and found that taxation was the most effective and cost-effective intervention in populations with a 5% or greater prevalence of heavy drinkers. The costs associated with this intervention included the cost of passing the legislation itself, and the cost of administering and enforcing the laws once they are passed. Effectiveness was assessed using disability-adjusted life-years (DALYs), a standard measure of global health impact that considers the impact of an intervention on healthy years of life lost as a result of either death or disability. For the Americas A region, consisting of the U.S., Canada, and Cuba, which is the region most relevant to this review, the intervention costs for current taxation were $482,956 (converted to 2007 dollars using the Consumer Price Index) per 1 million population per year, based on a 10-year implementation period and discounted at 3% per year to reflect the time value of money. The cost was assumed to stay the same when the tax was increased by 25% or 50%. Current taxes were estimated to prevent 1224 DALYs per 1 million population per year, yielding an average cost-effectiveness ratio for this intervention of approximately $395 per DALY averted. This is much less than the average annual income per capita in these three countries, a threshold for an intervention to be considered very cost effective that was proposed by the Commission on Macroeconomics and Health.[106] The DALYs averted increased to 1366 and 1489 per 1 million population per year when taxes were increased by 25% and 50%, respectively. Because these incremental DALYs averted could be achieved without any increase in costs, these increases in taxes improve cost-effectiveness estimates relative to the current tax scenario. To obtain country-specific estimates of the DALYs saved per country as a result of this intervention, the regional analysis needs to be adjusted using country-specific data. Such estimates are limited by the assumptions made and the data available.

Barriers to Implementation

The level of taxation of alcoholic beverages has economic effects on several groups, including federal,

state, and local governments; affected industry groups; and the general population of alcohol consumers. Whereas raising alcohol taxes may provide an important source of revenue for governments, such tax increases may be resisted by some industry groups and consumers. However, public support for increased alcohol taxes increases substantially when tax revenues are specifically directed to fund prevention and treatment programs instead of being used as an unrestricted source of general revenue.[107]

Other Benefits or Harms

In addition to the direct public health outcomes evaluated in this review, the primary benefit of increased alcohol excise taxes is that they can provide a source of revenue to support programs to prevent and treat alcohol problems. They also can provide some compensation for the societal costs associated with excessive alcohol consumption that are not borne by the drinker (i.e., "external" costs). Economic analyses suggest that alcohol taxes would need to be increased substantially to address adequately such external costs as crime, alcohol-related crashes, domestic violence, and productivity losses.[18,108]

A potential concern is that increases in alcohol taxes may have a greater proportional economic impact on people with lower incomes (i.e., alcohol taxes may be regressive). However, alcohol taxes constitute a minor proportion (i.e., <1%) of the tax burden of Americans, including those with low incomes. As such, concerns about the regressive nature of such taxes could be readily addressed by compensatory changes in other elements of the tax system. In addition, the amount of tax paid is directly related to the amount of alcohol consumed, and thus increases in alcohol excise taxes will be disproportionately paid by excessive drinkers, who also experience most of the alcohol-related harms and thus generate most alcohol-attributable economic costs. Furthermore, the beneficial economic results of reducing excessive alcohol consumption and related harms may also be disproportionately greater for people with low incomes. Lower-income people may be particularly vulnerable to the harmful consequences of excessive alcohol consumption—consumed by themselves or others—because of factors such as lower rates of health insurance coverage, which may result in lack of or incomplete treatment for alcohol-related illness or injuries. Increasing alcohol excise taxes could also directly benefit low-income populations if the revenue generated from these taxes is used to help improve the availability of healthcare services for uninsured and other vulnerable populations.

Summary

The reviewed studies provide consistent evidence that higher alcohol prices and alcohol taxes are associated with reductions in both excessive alcohol consumption and related, subsequent harms. Results were robust across different countries, time periods, study designs and analytic approaches, and outcomes. According to *Community Guide* rules of evidence,[19] these studies provide strong evidence that raising alcohol taxes is an effective strategy for reducing excessive alcohol consumption and related harms.

Most of the studies that were included in this review assessed the relationship between alcohol prices and the outcomes of interest using price elasticities. Alcohol-related harms that were well represented in the literature reviewed included alcohol-impaired driving, motor-vehicle crashes, various measures of violence, and liver cirrhosis. For the largest body of evidence in this review—that is, societal levels of alcohol consumption—the majority of estimates of price elasticity fell within the range of approximately −0.30 to −1.00, indicating that a 10% increase in alcohol prices would be expected to result in a 3% to 10% decrease in alcohol consumption. These results indicate that alcohol consumption is responsive to price, and suggest that the impact of a potential tax increase is likely to be proportional to its size. It would also be reasonable to expect that alcohol price elasticities may vary across population groups by age and disposable income, among other factors, but assessment of such group differences was not possible using results from the studies in this review.

Research Gaps

The volume and consistency of the evidence reviewed here suggests little need for additional research on the basic questions of whether changes in alcohol taxes and price affect excessive alcohol consumption and related harms. Nonetheless, studies published subsequent to the 2005 cutoff date for this review continue to indicate the public health benefits that accrue from increasing alcohol taxes. For example, a recent meta-analysis found very similar mean price elasticities for alcohol consumption as were found in this review.[109] Similarly, a recent study of alcohol-related disease mortality found that substantial alcohol tax increases in Alaska in 1983 and 2002 resulted in estimated reductions of 29% and 11%, respectively.[110]

However, additional research is needed to assess:

1. Whether changes in alcohol prices differentially affect drinking behavior and health outcomes for important subgroups of the population, such as underage young people.

2. The relative benefits of increasing taxes on all alcoholic beverages simultaneously, versus selectively increasing taxes on specific beverage types. This evaluation should be considered in light of known differences in the beverage preferences of binge drinkers, historic changes in tax rates across beverage types, and the effect of inflation on real tax rates by beverage type.
3. The impact of different approaches to taxing alcoholic beverages on excessive alcohol consumption and related harms. Specific emphasis should be placed on the impact of alcohol sales taxes, where taxes are calculated as a proportion of the total beverage price; the potential impact of standardizing alcohol taxes across beverage types based on alcohol content; and the potential impact of alcohol taxes levied by local governments on a per-drink basis in on-premise, retail alcohol outlets (i.e., tippler taxes).

Author affiliations are shown at the time the research was conducted. The names and affiliations of the Task Force members are listed at www.thecommunityguide.org.

The work of Briana Lawrence and Aneeqah Ferguson was supported with funds from the Oak Ridge Institute for Scientific Education (ORISE).

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. Mokdad AH, Stroup D, Marks J, Gerberding J. Actual causes of death in the U.S., 2000. JAMA 2004;291:1238–45.
2. NIAAA. Tenth Special Report to the U.S. Congress on alcohol and health. Bethesda MD: NIH, 2000.
3. CDC. Web-based Injury Statistics Query and Reporting System (WISQARS). 2008. www.cdc.gov/ncipc/wisqars.
4. Marshall A. Principles of economics. 8th ed. London: Macmillan, 1920.
5. Alcohol Epidemiology Program. Alcohol polices in the U.S.: highlights from the 50 States. Minneapolis: University of Minnesota, 2000.
6. Olson S, Gerstein DR. Alcohol in America: taking action to prevent abuse. Washington DC: National Academies Press, 1985.
7. Young DJ, Bielinska-Kwapisz A. Alcohol taxes and beverage prices. Natl Tax J 2002;55(1):57–73.
8. USDHHS. Healthy People 2010. 2nd ed. Washington DC: U.S. Government Printing Office, 2000.
9. Babor TF, Caetano R. Evidence-based alcohol policy in the Americas: strengths, weaknesses, and future challenges. Pan Am J Public Health 2005;18:327–37.
10. Chisolm D, Rehm J, Van Omeren M, et al. Reducing the global burden of hazardous alcohol use: a comparative cost-effectiveness analysis. J Stud Alcohol 2004;65:782–93.
11. Holder HD, Treno AJ. Moving toward a common evidence base for alcohol and other drug prevention policy. In: Stockwell T, Gruenwald PJ, Toumbourou JW, Loxley W, eds. Preventing harmful substance use: the evidence base for policy and practice. New York: John Wiley and Sons, 2005:351–66.
12. Toomey TL, Wagenaar AC. Policy options for prevention: the case of alcohol. J Public Health Policy 1999;20(2):192–213.
13. Babor TF, Caetano R, Casswell S, et al. Alcohol: no ordinary commodity: research and public policy. New York: Oxford University Press, 2003.
14. Chaloupka FJ, Grossman M, Saffer H. The effects of price on alcohol consumption and alcohol-related problems. Alcohol Res Health 2002;26:22–34.
15. IOM. Reducing underage drinking: a collective responsibility. Washington DC: National Academies Press, 2004.
16. Surgeon General's workshop on drunk driving: proceedings. Rockville MD: USDHHS, 1989.
17. Cook PJ, Moore MJ. The economics of alcohol abuse and alcohol-control policies. Health Aff (Millwood) 2002;21(2):120–33.
18. Cook PJ. Paying the tab: the costs and benefits of alcohol control. Princeton NJ: Princeton University Press, 2007.
19. Briss PA, Zaza S, Pappaioanou M, et al. Developing an evidence-based Guide to Community Preventive Services—methods. The Task Force on Community Preventive Services. Am J Prev Med 2000;18(1S):35–43.
20. Task Force on Community Preventive Services. Increasing alcoholic beverage taxes is recommended to reduce excessive alcohol consumption and related harms. Am J Prev Med 2010;38(2):230–2.
21. 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;33:188–93.
22. Cook PJ, Skog O-J. [Discussion of] "Alcool, alcoolisme, alcoolisation" by S. Ledermann. Alcohol Health Res World 1995;19:30–2.
23. Skog O-J. The collectivity of drinking cultures: a theory of the distribution of alcohol consumption. Br J Addict 1985;80:83–99.
24. Dee TS. State alcohol policies, teen drinking and traffic fatalities. J Public Econ 1999;72(2):289–315.
25. Cook PJ, Tauchen G. The effect of liquor taxes on heavy drinking. Bell J Econ 1982;13(2):379–90.
26. Chaloupka FJ, Saffer H, Grossman M. Alcohol-control policies and motor-vehicle fatalities. J Legal Stud 1993;22:161–86.
27. Blake D, Nied A. The demand for alcohol in the United Kingdom. Appl Econ 1997;29:1655–72.
28. Heien D, Pompelli G. Stress, ethnic, and distribution factors in a dichotomous response model of alcohol abuse. J Stud Alcohol 1987;48(5):450–5.
29. Ruhm CJ. Economic conditions and alcohol problems. J Health Econ 1995;14:583–603.
30. Ruhm CJ. Alcohol policies and highway vehicle fatalities. J Health Econ 1996;15:435–54.
31. Cook PJ, Ostermann J, Sloan FA. Are alcohol excise taxes good for us? Short- and long-term effects on mortality rates.

Working Paper No. 11138. Cambridge MA: National Bureau of Economic Research, 2005.
32. Beard TR, Gant PA, Saba RP. Border-crossing sales, tax avoidance, and state tax policies: an application to alcohol. South Econ J 1997;64(1):293–306.
33. Baltagi BH, Goel RK. Quasi-experimental price elasticity of liquor demand in the U.S.: 1960–83. Am J Agric Econ 1990;72(2):451–4.
34. Simon JL. Price elasticity of liquor in the US and a simple method of determination. Econometrica 1966;34(1):193–205.
35. Decker SL, Schwartz AE. Cigarettes and alcohol: substitutes or complements. Working Paper No. 7535. Cambridge MA: National Bureau of Economic Research, 2000.
36. Levy D, Sheflin N. New evidence on controlling alcohol use through price. J Stud Alcohol 1983;44(6):929–37.
37. Johnson JA, Oksanen EH. Socioeconomic determinants of the consumption of alcoholic beverages. Appl Econ 1974; 6(4):293–301.
38. Goel RK, Morey MJ. The interdependence of cigarette and liquor demand. South Econ J 1995;62(2):451–9.
39. Hoadley JF, Fuchs BC, Holder HD. The effect of alcohol beverage restriction on consumption: a 25-year longitudinal analysis. Am J Drug Alcohol Abuse 1984;10(3):375–401.
40. Lee B, Tremblay VJ. Advertising and the US market demand for beer. Appl Econ 1992;24(1):69–76.
41. Ornstein SI, Hanssens DM. Alcohol control laws and the consumption of distilled spirits and beer. J Consum Res 1985;12(2):200–13.
42. Saffer H, Grossman M. Drinking age laws and highway mortality rates: cause and effect. Econ Inq 1987;25(3):403–17.
43. Wilkinson JT. Reducing drunken driving: which policies are most effective? South Econ J 1987;54:322–34.
44. Adrian M, Ferguson BS, Her M. Can alcohol price policies be used to reduce drunk driving? Evidence from Canada. Subst Use Misuse 2001;36(13):1923–57.
45. Wette HC, Zhang JF, Berg RJ, Casswell S. Effect of prices on alcohol consumption in New Zealand 1983–1991. Drug Alcohol Rev 1993;12(2):151–8.
46. Grossman M, Chaloupka FJ, Sirtalan I. An empirical analysis of alcohol addiction: results from the monitoring the future panels. Econ Inq 1998;36(1):39–48.
47. Hogarty TF, Elzinga KG. The demand for beer. Rev Econ Stat 1972;54(2):195–8.
48. Mayo JR. An estimate of U.S. demand for alcoholic beverages, 1986–92. Penn Econ Rev 2000;9(1):1–4.
49. Bourgeois JC, Barnes JG. Does advertising increase alcohol consumption? J Advert Res 1979;19(4):19–29.
50. Speer PW, Gorman DM, Labouvie EW, Ontkush MJ. Violent crime and alcohol availability: relationships in an urban community. J Public Health Policy 1998;19(3):303–18.
51. Farrell S, Manning WG, Finch MD. Alcohol dependence and the price of alcoholic beverages. J Health Econ 2003; 22(1):117–47.
52. Gruenewald PJ, Madden P, Janes K. Alcohol availability and the formal power and resources of state alcohol beverage control agencies. Alcohol Clin Exp Res 1992;16(3):591–7.
53. Manning WG, Blumberg L, Moulton LH. The demand for alcohol: the differential response to price. J Health Econ 1995;14(2):123–48.
54. Kuo M, Heeb JL, Gmel G, Rehm J. Does price matter? The effect of decreased price on spirits consumption in Switzerland. Alcohol Clin Exp Res 2003;27(4):720–5.
55. Gius MP. The effect of taxes on alcoholic consumption: an individual level of analysis with a correction for aggregate public policy variables. Penn Econ Rev 2002;11(1):76–93.
56. Laixuthai A, Chaloupka FJ. Youth alcohol use and public policy. Contemp Policy Issues 1993;11(4):70–81.
57. Heien DM, Pompelli G. The demand for alcoholic beverages: economic and demographic effects. South Econ J 1989; 55(3):759–70.
58. Coate D, Grossman M. Effects of alcoholic beverage prices and legal drinking ages on youth alcohol use. J Law Econ 1988;31(1):145–71.
59. Heeb J-L, Gmel G, Zurbrugg C, Kuo M, Rehm J. Changes in alcohol consumption following a reduction in the price of spirits: a natural experiment in Switzerland. Addiction 2003;98(10):1433–46.
60. Pacula RL. Does increasing the beer tax reduce marijuana consumption? J Health Econ 1998;17:557–85.
61. Kenkel DS. Drinking, driving, and deterrence: the effectiveness and social costs of alternative policies. J Law Econ 1993;36:877–911.
62. Grossman M, Coate D, Arluck GM. Price sensitivity of alcoholic beverages in the U.S.: youth alcohol consumption. In: Holder H, ed. Control issues in alcohol abuse prevention: strategies for states and communities. Greenwich CT: JAI Press, 1987;169–98.
63. Nelson J. State monopolies and alcoholic beverage consumption. J Regul Econ 1990;2:83–98.
64. Chaloupka FJ, Wechsler H. Binge drinking in college: the impact of price, availability, and alcohol control policies. Contemp Econ Policy 1996;14(4):112–24.
65. Yen ST. Cross-section estimation of US demand for alcoholic beverage. Appl Econ 1994;26(4):381–92.
66. Mullahy J, Sindelar JL. Do drinkers know when to say when? An empirical analysis of drunk driving. Econ Inq 1994; 32(3):383–94.
67. Uri ND. The demand for beverages and interbeverage substitution in the U.S. Bull Econ Res 1986;38(1):77–85.
68. Cook PJ, Moore MJ. This tax's for you: the case for higher beer taxes. Natl Tax J 1994;47(3):559–73.
69. Markowitz S, Grossman M. Alcohol regulation and domestic violence towards children. Contemp Econ Policy 1998;16(3): 309–20.
70. Sutton M, Godfrey C. A grouped data regression approach to estimating economic and social influences on individual drinking behaviour. Health Econ 1995;4(3):237–47.
71. Duffy M. Influence of prices, consumer incomes, and advertising upon the demand for alcoholic drink in the United Kingdom: an econometric study. Alcohol Alcohol 1981; 16(4):200–8.
72. Sloan FA, Reilly BA, Schenzler C. Effects of prices, civil and criminal sanctions, and law enforcement on alcohol-related mortality. J Stud Alcohol 1994;55:454–65.
73. Johnson JA, Oksanen EH, Veall MR, Fretz D. Short-run and long-run elasticities for Canadian consumption of alcoholic beverages: an error-correction mechanism/cointegration approach. Rev Econ Stat 1992;74(1):64–74.

74. Nelson J, Moran J. Advertising and US alcohol beverage system demand: system-wide estimates. Appl Econ 1995; 12:1225–36.
75. Treno AJ, Parker RN, Holder HD. Understanding U.S. alcohol consumption with social and economic factors: a multivariate time series analysis, 1950–1986. J Stud Alcohol 1993;54:146–56.
76. Gray D, Chikritzhs T, Stockwell T. The Northern Territory's cask wine levy: health and taxation policy implications. Aust N Z J Public Health 1999;23(6):651–3.
77. Zhang JF, Casswell S. The effects of real price and a change in the distribution system on alcohol consumption. Drug Alcohol Rev 1999;18:371–8.
78. Clements KW, Selvanathan S. The economic determinants of alcohol consumption. Aust J Agric Resour Econ 1991; 35(2):209–31.
79. Duffy M. The demand for alcoholic drink in the United Kingdom. Appl Econ 1983;15(1):125–40.
80. Labys W. An international comparison of price and income elasticities for wine consumption. Aust J Agric Resour Econ 1976;20(1):33–6.
81. Nelson J. Economic and demographic factors in U.S. alcohol demand: a growth-accounting analysis. Empir Econ 1997; 22:83–102.
82. Selvanathan EA. Alcohol consumption in the UK, 1955–85: a system-wide analysis. Appl Econ 1988;20(2):1071–86.
83. Rush B, Steinberg M, Brook R. Relationships among alcohol availability, alcohol consumption and alcohol-related damage in the province of Ontario and the State of Michigan. Adv Alcohol Subst Abuse 1986;5(4):33–45.
84. Young C, Bielinska-Kwapisz A. Alcohol consumption, beverage prices and measurement error. J Stud Alcohol 2003; 64:235–8.
85. Markowitz S, Grossman M. The effects of beer taxes on physical child abuse. J Health Econ 2000;19:271–82.
86. Evans WN, Neville D, Graham JD. General deterrence of drunk driving: evaluation of recent American policies. Risk Anal 1991;11:279–89.
87. Chesson H, Harrison P, Kassler WJ. Sex under the influence: the effect of alcohol policy on sexually transmitted disease rates in the U.S. J Law Econ 2000;43:215–37.
88. Brinkley G. The causal relationship between socioeconomic factors and alcohol consumption: a Granger-causality time series analysis, 1950–1993. J Stud Alcohol 1999;60(6): 759–68.
89. Niskanen WA. Taxation and the demand for alcoholic beverages. Santa Monica CA: Rand Corp, 1960.
90. 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.
91. Kendell RE, Ritson B. Effect of economic changes on Scottish drinking habits 1978–82. Br J Addict 1983;78:365–79.
92. Adrian M, Ferguson BS. Demand for domestic and imported alcohol in Canada. Appl Econ 1987;19(4):531–40.
93. Cook PJ. The effect of liquor taxes on drinking, cirrhosis, and auto accidents. In: Moore MH, Gerstein D, eds. Alcohol and public policy: beyond the shadow of prohibition. Washington DC: National Academies Press, 1981:255–85.
94. Duffy M. Advertising and the inter-product distribution of demand. Eur Econ Rev 1987;31:1051–70.
95. Jones AM. A systems approach to the demand for alcohol and tobacco. Bull Econ Res 1989;41(2):85–101.
96. Kenkel DS. New estimates of the optimal tax on alcohol. Econ Inq 1996;34(2):296–319.
97. Leskinen E, Terasvirta T. Forecasting the consumption of alcoholic beverages in Finland. Eur Econ Rev 1976;8:349–69.
98. Mast BD, Benson BL, Rasmussen DW. Beer taxation and alcohol-related traffic fatalities. South Econ J 1999;66(2): 214–49.
99. Ponicki W, Holder HD, Gruenewald PRA. Altering alcohol price by ethanol content: results from a Swedish tax policy in 1992. Addiction 1997;92(7):859–70.
100. 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;32:723–33.
101. Cook PJ, Moore MJ. Economic perspectives on reducing alcohol-related violence. Alcohol and interpersonal violence: fostering multidisciplinary perspectives. NIAAA Research Monograph 24. Rockville MD: NIAAA, 1993:193–212.
102. Stout EM, Sloan FA, Liang L, Davies HH. Reducing harmful alcohol-related behaviors: effective regulatory methods. J Stud Alcohol 2000;61(3):402–12.
103. Saffer H, Grossman M. Beer taxes, the legal drinking age, and youth motor vehicle fatalities. J Legal Stud 1987;16(June): 351–73.
104. Chaloupka FJ. Effects of price on alcohol-related problems. Alcohol Health Res World 1993;17(1):46–53.
105. Miller TR, Levy, DT. Cost-outcome analysis in injury prevention and control: eighty-four recent estimates for the U.S. Med Care 2000;38(6):562–82.
106. WHO. Macroeconomics and health: investing in health for economic development. Final report of the Commission on Macroeconomics and Health. Geneva, Switzerland: WHO, 2001.
107. Wagenaar AC, Harwood EH, Toomey TL, Denk CE, Zander KM. Public opinion on alcohol policies in the U.S.: results from a national survey. J Public Health Policy 2000;21: 303–27.
108. Richardson J, Crowley S. Optimum alcohol taxation: balancing consumption and external costs. Health Econ 1994; 3(2):73–87.
109. Wagenaar AC, Salois MJ, Komro KA. Effects of beverage alcohol price and tax levels on drinking: a meta-analysis of 1003 estimates from 112 studies. Addiction 2009;104:179–90.
110. Wagenaar AC, Maldonado-Molina MM, Wagenaar BH. Effects of alcohol tax increases on alcohol-related disease mortality in Alaska: time–series analyses from 1976 to 2004. Am J Public Health 2009;99:1464–70.

nature genetics

Thirty new loci for age at menarche identified by a meta-analysis of genome-wide association studies

To identify loci for age at menarche, we performed a meta-analysis of 32 genome-wide association studies in 87,802 women of European descent, with replication in up to 14,731 women. In addition to the known loci at *LIN28B* ($P = 5.4 \times 10^{-60}$) and 9q31.2 ($P = 2.2 \times 10^{-33}$), we identified 30 new menarche loci (all $P < 5 \times 10^{-8}$) and found suggestive evidence for a further 10 loci ($P < 1.9 \times 10^{-6}$). The new loci included four previously associated with body mass index (in or near *FTO*, *SEC16B*, *TRA2B* and *TMEM18*), three in or near other genes implicated in energy homeostasis (*BSX*, *CRTC1* and *MCHR2*) and three in or near genes implicated in hormonal regulation (*INHBA*, *PCSK2* and *RXRG*). Ingenuity and gene-set enrichment pathway analyses identified coenzyme A and fatty acid biosynthesis as biological processes related to menarche timing.

Menarche, the onset of first menstruation in girls, indicates the attainment of reproductive capacity and is a widely used marker of pubertal timing. Age at menarche varies widely between girls and is highly dependent on nutritional status[1]. Early menarche is associated with several adverse health outcomes, including breast cancer[2], endometrial cancer[3], obesity[4], type 2 diabetes[5] and cardiovascular disease[6], as well as shorter adult stature[4]. Studies of twins and extended families, although largely performed in populations free of nutritional deprivation, estimate that around 50% of the variance in menarche timing is attributable to genetic factors in such settings[7].

Recently, common variants in *LIN28B* were associated with age at menarche in four independent genome-wide association studies (GWAS)[8–11]. *LIN28B* is a human homolog of *lin-28* in *Caenorhabditis elegans*, which controls the rate of progression from larval stages to adult cuticle formation, indicating the possible conservation of specific micro-RNA regulatory mechanisms involved in developmental timing[9]. A second menarche locus was identified in an intergenic region at 9q31.2[8,10]. These two loci together explained only 0.6% of the variance in age at menarche[8]. We anticipated that a much larger GWAS would substantially increase the yield of loci associated with age at menarche.

Here we report a much expanded meta-analysis of GWAS for age at menarche. By combining data from the previous studies[8–11], plus several further studies to form the ReproGen Consortium, we identified at least 30 previously unidentified loci associated with age at menarche at genome-wide significance levels. Our findings show a close link between the genetic regulation of energy homeostasis and pubertal timing and suggest the presence of other diverse pathways.

RESULTS

Genome-wide association for age at menarche

This expanded GWAS includes data from 32 cohorts of European ancestry ($N = 87{,}802$). In most studies, age at menarche was determined by self recall, and the mean age at menarche in individual studies ranged from 12.4 to 13.6 years, excluding individuals with menarche <9 years and >17 years (Online Methods, **Supplementary Table 1** and **Supplementary Note**). Genome-wide SNP genotyping was performed using a variety of different platforms (**Supplementary Table 2** and **Supplementary Note**). Therefore, after applying standard quality control measures, we imputed the genotypes for ~2.5 million autosomal SNPs in the HapMap European CEU sample using Build 35 or 36 to allow inverse variance meta-analysis of additive genetic association results from each study. We also meta-analyzed results from X-chromosome SNPs in studies which had this data available ($N = 52{,}781$). Test statistics from each cohort were adjusted using genomic control to avoid inflation of results due to population stratification.

There was strong deviation from the uniform distribution of *P* values expected under the null hypothesis (**Supplementary Fig. 1**). This deviation was attenuated, but persisted, following removal of those signals associated with the two previously identified loci. In total, 945 SNPs representing 45 loci ($r^2 < 0.05$ based on HapMap in a 750-kb region) were associated with age at menarche at genome-wide significance levels ($P < 5 \times 10^{-8}$) (**Fig. 1** and **Supplementary Fig. 2**). None of these loci were located on the X chromosome. These 45 loci included three apparent second signals (defined as two genome-wide significant SNPs in low linkage disequilibrium (LD) ($r^2 < 0.05$) in the same 750-kb region) at 2q33.1, 6q21 and 14q32.2. The second signal at 6q21 (rs314279) had a low minor allele frequency (MAF = 6%) and was not present in many studies. We therefore genotyped this SNP *de novo* in the InCHIANTI cohort and found it was in LD with the top chromosome 6 signal (rs7759938, $r^2 = 0.3$). In HapMap, the r^2 between the two chromosome 6 SNPs was 0.015, but the *D* was 1.0. To verify the independence of additional loci, we performed a conditional analysis and a meta-analysis of all 32 studies using the top SNPs at all the 42 genome-wide significant regions as covariates (in addition to birth year). In these conditional analyses, the possible second signals on chromosomes 2 and 14 showed strong but not genome-wide significant associations with age at menarche ($P < 7.1 \times 10^{-6}$), suggestive of, but not confirming, second independent signals in these two regions (**Fig. 1** and **Supplementary Table 3**).

A full list of authors and affiliations appear at the end of the paper.

Received 21 May; accepted 19 October; published online 21 November 2010; doi:10.1038/ng.714

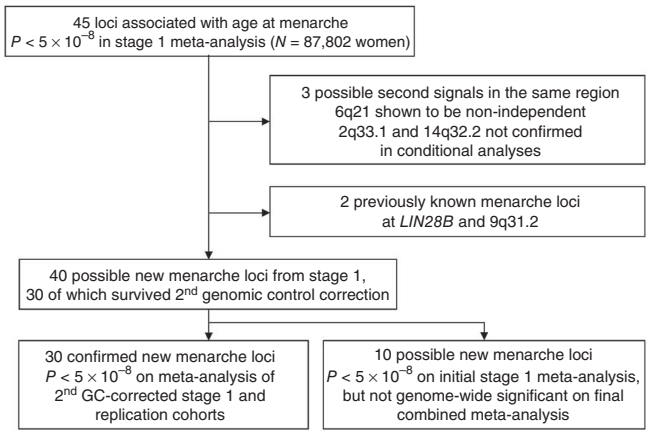


Figure 1 Flow diagram of the discovery and confirmation of new loci for age at menarche. GC, genomic control.

The two most significant loci for age at menarche confirmed the previously reported associations at *LIN28B* (rs7759938, $P = 1.6 \times 10^{-58}$) and 9q31.2 (rs2090409, $P = 4.4 \times 10^{-33}$) (**Table 1** and **Supplementary Fig. 3**). In addition, there were genome-wide significant signals for a further 40 possible previously unidentified loci, of which 30 survived a second more stringent correction for the overall genomic control in the stage 1 cohorts ($\lambda = 1.173$) (**Table 1**, **Fig. 1** and **Supplementary Fig. 3**).

Replication studies

We sought confirmation of the 40 possible new menarche loci in up to 14,731 women from 16 additional studies with *in silico* GWAS data and new genotyping data from one cohort (**Supplementary Tables 4** and **5**). This replication sample was substantially smaller than our stage 1 sample and was therefore underpowered to confirm individual SNP associations (**Supplementary Fig. 4**). Nonetheless, 37 of the 40 possible newly associated loci showed directionally consistent associations in both stages (**Table 1**; binomial sign test $P = 9.7 \times 10^{-9}$). A combined meta-analysis of the more stringent second genomic control–corrected stage 1 results and replication cohorts gave confirmatory evidence for 30 new menarche loci, leaving 10 unconfirmed possible menarche loci (**Table 1** and **Fig. 1**).

Based on the combined stage 1 and replication results, the estimated magnitudes of per-allele effects for the new menarche loci ranged from 4.5 to 2.1 weeks per allele (**Table 1**) and had an inverse relationship with MAF (**Supplementary Fig. 5**). Among the four largest *in silico* replication cohorts (each comprising >800 women), the variance in age at menarche explained by all 42 known, confirmed and possible new menarche loci ranged from 3.6% to 6.1% (**Supplementary Table 6**).

Candidate genes at new loci

The strongest new menarche signal was for rs1079866 (3.9 weeks per minor allele; 95% CI 2.9–5.0, $P = 5.5 \times 10^{-14}$) located approximately 250 kb downstream of *INHBA*, which encodes the protein subunit Inhibin beta A. Heterodimers of Inhibin beta A and the Inhibin alpha subunit form the female reproductive hormone Inhibin A[12]. Inhibin A, produced by granulosa cells in the ovary, increases dramatically during pubertal development in girls[13,14] and is involved in negative feedback regulation by inhibiting production of follicle stimulating hormone by the pituitary and secretion of gonadotrophin releasing hormone from the hypothalamus[15]. Conversely, homodimers of Inhibin beta A form the hormone Activin A, which stimulates pituitary follicle stimulating hormone production and also exhibits a wide range of biological activities, including the regulation of cellular proliferation and differentiation[16].

The second strongest new signal was for rs466639 ($P = 1.3 \times 10^{-13}$); this SNP is intronic in *RXRG*, which encodes retinoid X receptor gamma, a nuclear receptor that forms dimers with the receptors for retinoic acid, thyroid hormone and vitamin D, increasing both DNA binding and transcriptional function on their respective response elements[17].

Four new loci for menarche were previously identified by GWAS for adult body mass index (BMI)[18–20]: rs9939609 (in or near *FTO*, $P = 3.1 \times 10^{-8}$), rs633715 (*SEC16B*, $P = 2.1 \times 10^{-8}$), rs2002675 (*TRA2B* and *ETV5*, $P = 1.2 \times 10^{-9}$) and rs2947411 (*TMEM18*, $P = 1.7 \times 10^{-8}$). Apart from rs2002675, these menarche signals were either identical to or in tight LD ($r^2 > 0.9$) with those BMI loci, and in all cases, the BMI-increasing allele was associated with earlier menarche. Variants at these four loci have also been associated with childhood BMI[18–20], and these findings support a likely causal effect of childhood BMI on earlier pubertal timing.

Three new menarche loci were found in or near further genes implicated in the regulation of energy homeostasis and body weight in animal models: rs6589964 ($P = 1.9 \times 10^{-12}$) lies ~18 kb from *BSX*, rs10423674 ($P = 5.9 \times 10^{-9}$) is intronic in *CRTC1*, and rs4840046 ($P = 2.4 \times 10^{-8}$) lies ~160 kb from *MCHR2*. *BSX* encodes a DNA-binding protein and transcriptional activator. In mouse, *Bsx* is expressed specifically in the pineal gland, telencephalic septum, hypothalamic pre-mammillary body and arcuate nucleus and is necessary for postnatal growth, locomotory behavior, expression of the genes *Npy* and *Agrp*, and for the hyperphagic phenotype in leptin deficiency[21]. *CRTC1* encodes the CREB-regulated transcription coactivator 1, an activator of cellular gene expression. *Crtc1*$^{-/-}$ mice are hyperphagic, obese and infertile, and *Crtc1*$^{-/-}$ females have low circulating luteinizing hormone levels[22]. Leptin potentiates the effects of *Crtc1* transcriptional activity, and *Crtc1* overexpression in hypothalamic cells increases expression of Kisspeptin, which in turn activates secretion of the gonadotrophin releasing hormone. *MCHR2* encodes the melanin concentrating hormone receptor 2, an orphan G protein–coupled receptor which shows high affinity binding to the hypothalamic neuropeptide melanin-concentrating hormone (MCH), which regulates nutrient intake and energy homeostasis through *MCHR1*[23]. Furthermore, MCH directly inhibits gonadotrophin releasing hormone neurons and thereby links energy balance to reproduction[24].

rs852069 ($P = 3.3 \times 10^{-8}$) lies ~84 kb from *PCSK2*, which encodes proprotein convertase subtilisin/kexin type 2, an enzyme that cleaves latent precursor proteins, such as proinsulin and proopiomelanocortin, into their biologically active products. Although rare deleterious mutations and common variants in *PCSK1* are known to influence obesity risk, it is notable that *PCSK2* differs from *PCSK1* in that it additionally cleaves pro-luteinizing hormone-releasing hormone and could therefore have a more direct influence on the reproductive hormone axis.

Pathway analyses

Remaining new menarche loci were found in or near genes that are involved in a seemingly diverse range of biological functions (**Supplementary Table 7**). We used ingenuity pathway analysis (IPA) to identify potential biological pathways common to these identified loci. Based on direct interactions only, we identified two functional networks containing 16 and 11 genes, respectively, of those genes nearest to the new menarche loci (**Supplementary Fig. 6**). Network 1,

related to 'gene expression, cellular growth and proliferation, and cellular function and maintenance', covers a wide and nonspecific range of biological pathways. Functions in network 2 relate to 'lipid metabolism, small molecule biochemistry and molecular transport' (**Supplementary Table 8**). Central to network 2 are *RXRG* and several genes involved in fatty acid biosynthesis, including several fatty acid–binding proteins and *ACSL1*, which encodes an enzyme that converts free long-chain fatty acids into fatty acyl-CoA esters.

To identify potential further biological pathways that influence menarche timing, we used a gene set enrichment analysis (GSEA)

Table 1 Stage 1 and replication results for 42 known, confirmed or possible new loci for age at menarche

							Stage 1			Replication		Stage 1 and replication			
SNP	Nearest gene(s)	Distance from gene (kb)	Chr.	Position (Build 36)	MAF[a]	Alleles[b]	P_{het}[c]	P[d]	P[e 2-GC]	n	P[f]	β[g]	s.e.	Direction[h]	P[i]
Previous menarche loci															
rs7759938[j]	*LIN28B*	~26 kb	6	105,485,647	0.32	C/T	0.04	1.6×10^{-58}	4.3×10^{-50}	14,185	4.6×10^{-11}	6.4	0.4	+/+	5.4×10^{-60}
rs2090409	*TMEM38B*	~400 kb	9	108,006,909	0.31	A/C	0.05	4.4×10^{-33}	2.3×10^{-28}	14,708	2.7×10^{-6}	–4.7	0.4	–/–	2.2×10^{-33}
30 novel menarche loci															
rs1079866	*INHBA*	~250 kb	7	41,436,618	0.15	G/C	0.81	1.9×10^{-16}	2.7×10^{-14}	14,731	1.9×10^{-1}	3.9	0.5	+/+	5.5×10^{-14}
rs466639	*RXRG*	Intronic	1	163,661,506	0.13	T/C	0.80	7.8×10^{-15}	8.9×10^{-13}	14,279	3.1×10^{-2}	–4.2	0.6	–/–	1.3×10^{-13}
rs6438424	*3q13.32*	Intergenic	3	119,057,512	0.50	A/C	0.99	8.4×10^{-14}	4.6×10^{-12}	8,634	6.7×10^{-3}	–2.7	0.4	–/–	1.4×10^{-13}
rs1398217	*FUSSEL18*	Intronic	18	43,006,236	0.43	G/C	0.33	5.7×10^{-13}	2.5×10^{-11}	14,344	2.3×10^{-3}	–2.7	0.4	–/–	2.3×10^{-13}
rs12617311	*PLCL1*	~195 kb	2	199,340,810	0.32	A/G	0.90	2.6×10^{-13}	1.2×10^{-11}	14,007	1.1×10^{-2}	–3.0	0.4	–/–	6.0×10^{-13}
rs9635759	*CA10*	~94 kb	17	46,968,784	0.32	A/G	0.43	2.0×10^{-13}	1.5×10^{-11}	14,002	1.1×10^{-2}	3.0	0.4	+/+	7.3×10^{-13}
rs6589964	*BSX*	~18 kb	11	122,375,893	0.48	A/C	0.89	8.8×10^{-14}	4.3×10^{-12}	13,754	8.3×10^{-2}	–2.7	0.4	–/–	1.9×10^{-12}
rs10980926	*ZNF483*	Intronic	9	113,333,455	0.36	A/G	0.65	2.2×10^{-13}	9.2×10^{-12}	14,227	3.8×10^{-1}	2.5	0.4	+/+	4.2×10^{-11}
rs17268785	*CCDC85A*	Intronic	2	56,445,587	0.17	G/A	0.82	6.8×10^{-11}	2.0×10^{-9}	14,233	1.5×10^{-2}	3.2	0.5	+/+	9.7×10^{-11}
rs13187289	*PHF15*	~12 kb	5	133,877,076	0.20	G/C	0.99	2.0×10^{-10}	3.6×10^{-9}	14,303	1.4×10^{-2}	3.0	0.5	+/+	1.9×10^{-10}
rs7642134	*VGLL3*	~70 kb	3	86,999,572	0.38	A/G	0.65	2.3×10^{-9}	4.3×10^{-8}	14,205	2.1×10^{-3}	–2.4	0.4	–/–	3.5×10^{-10}
rs17188434	*NR4A2*	~84 kb	2	156,805,022	0.07	C/T	0.59	3.4×10^{-11}	9.1×10^{-10}	14,356	2.2×10^{-1}	–4.5	0.7	–/–	1.1×10^{-9}
rs2002675	*TRA2B, ETV5*	~4 kb, ~135 kb	3	187,112,262	0.42	G/A	0.94	3.9×10^{-9}	4.7×10^{-8}	14,334	6.6×10^{-3}	2.2	0.4	+/+	1.2×10^{-9}
rs7821178	*PXMP3*	~181 kb	8	78,256,392	0.34	A/C	0.38	6.7×10^{-10}	1.2×10^{-8}	14,151	8.0×10^{-2}	–2.4	0.4	–/–	3.0×10^{-9}
rs1659127	*MKL2*	~28 kb	16	14,295,806	0.34	A/G	0.19	3.0×10^{-9}	4.5×10^{-8}	14,021	2.5×10^{-2}	2.4	0.4	+/+	4.0×10^{-9}
rs10423674	*CRTC1*	Intronic	19	18,678,903	0.35	A/C	0.79	1.1×10^{-9}	1.7×10^{-8}	13,543	1.1×10^{-1}	2.3	0.4	+/+	5.9×10^{-9}
rs10899489	*GAB2*	Intronic	11	77,773,021	0.15	A/C	0.16	2.4×10^{-10}	4.7×10^{-9}	14,201	2.5×10^{-1}	3.1	0.5	+/+	8.1×10^{-9}
rs6575793	*BEGAIN*	Intronic	14	100,101,970	0.42	C/T	0.51	1.7×10^{-10}	3.7×10^{-9}	13,899	4.6×10^{-1}	2.3	0.4	+/+	1.2×10^{-8}
rs4929923	*TRIM66*	3′UTR	11	8,595,776	0.36	T/C	0.99	2.4×10^{-8}	2.2×10^{-7}	8,510	1.6×10^{-2}	2.3	0.4	+/+	1.2×10^{-8}
rs6439371	*TMEM108, NPHP3*	~146 kb, ~170 kb	3	134,093,442	0.34	G/A	0.35	1.5×10^{-8}	1.6×10^{-7}	8,581	3.0×10^{-2}	2.3	0.4	+/+	1.3×10^{-8}
rs900145	*ARNTL*	~5 kb	11	13,250,481	0.30	C/T	0.35	7.7×10^{-9}	1.1×10^{-7}	8,649	$6.5 \times 10^{--2}$	2.3	0.4	+/+	1.6×10^{-8}
rs6762477	*RBM6*	Intronic	3	50,068,213	0.44	G/A	0.22	1.4×10^{-9}	2.4×10^{-8}	12,447	1.5×10^{-1}	2.5	0.4	+/+	1.6×10^{-8}
rs2947411	*TMEM18*	~53 kb	2	604,168	0.17	A/G	0.27	2.1×10^{-8}	2.6×10^{-7}	8,657	1.9×10^{-2}	2.8	0.5	+/+	1.7×10^{-8}
rs1361108	*C6orf173, TRMT11*	~98 kb, ~407 kb	6	126,809,293	0.46	T/C	0.76	2.6×10^{-9}	3.0×10^{-8}	14,126	6.0×10^{-2}	–2.1	0.4	–/–	1.7×10^{-8}
rs1364063	*NFAT5*	~10 kb	16	68,146,073	0.43	C/T	0.05	4.4×10^{-8}	4.8×10^{-7}	8,669	7.1×10^{-3}	2.1	0.4	+/+	1.8×10^{-8}
rs633715	*SEC16B*	~44 kb	1	176,119,203	0.20	C/T	0.45	1.5×10^{-9}	2.3×10^{-8}	14,274	1.9×10^{-1}	–2.6	0.5	–/–	2.1×10^{-8}
rs4840086	*PRDM13, MCHR2*	~145 kb, ~160 kb	6	100,315,159	0.42	G/A	0.98	8.2×10^{-9}	1.2×10^{-7}	8,669	7.5×10^{-2}	–2.1	0.4	–/–	2.4×10^{-8}
rs7617480	*KLHDC8B*	Intronic	3	49,185,736	0.22	A/C	0.64	1.8×10^{-9}	2.7×10^{-8}	14,341	2.4×10^{-1}	2.4	0.4	+/+	2.8×10^{-8}
rs9939609	*FTO*	Intronic	16	52,378,028	0.40	A/T	0.17	3.3×10^{-11}	1.1×10^{-9}	8,665	5.3×10^{-1}	–2.1	0.4	–/+	3.1×10^{-8}
rs852069	*PCSK2*	~84 kb	20	17,070,593	0.37	A/G	0.47	1.1×10^{-9}	2.0×10^{-8}	14,306	3.3×10^{-1}	–2.1	0.4	–/–	3.3×10^{-8}
10 possible menarche loci[k]															
rs757647	*KDM3B*	Intronic	5	137,735,214	0.22	A/G	0.23	1.4×10^{-9}	2.0×10^{-8}	14,326	4.4×10^{-1}	–2.4	0.4	–/–	5.4×10^{-8}
rs9555810	*C13orf16, ARHGEF7*	~185 kb, ~223 kb	13	110,979,438	0.28	G/C	0.68	6.7×10^{-10}	1.4×10^{-8}	14,266	4.9×10^{-1}	2.3	0.4	+/+	5.6×10^{-8}
rs16938437	*PHF21A*	Intronic	11	46,009,151	0.09	T/C	0.32	1.4×10^{-9}	2.2×10^{-8}	14,330	3.8×10^{-1}	–3.7	0.7	–/–	5.9×10^{-8}
rs2687729	*EEFSEC*	Intronic	3	129,377,916	0.27	G/A	0.36	1.0×10^{-8}	1.4×10^{-7}	8,669	3.2×10^{-1}	2.3	0.4	+/+	1.3×10^{-7}
rs1862471	*OLFM2*	Intronic	19	9,861,322	0.47	G/C	0.17	4.6×10^{-10}	8.3×10^{-9}	13,470	9.4×10^{-1}	2.0	0.4	+/–	1.5×10^{-7}
rs12472911	*LRP1B*	Intronic	2	141,944,979	0.20	C/T	0.65	3.9×10^{-8}	3.9×10^{-7}	8,585	1.4×10^{-1}	2.5	0.5	+/+	1.5×10^{-7}
rs3914188	*ECE2*	3′ UTR	3	185,492,742	0.27	G/C	0.54	2.3×10^{-9}	3.2×10^{-8}	14,085	7.9×10^{-1}	–2.2	0.4	–/–	2.6×10^{-7}
rs2243803	*SLC14A2*	~238 kb	18	41,210,670	0.40	A/T	0.89	2.8×10^{-8}	3.3×10^{-7}	8,659	3.9×10^{-1}	2.0	0.4	+/+	3.4×10^{-7}
rs3743266	*RORA*	3′ UTR	15	58,568,805	0.32	C/T	0.24	2.6×10^{-8}	2.9×10^{-7}	8,666	7.8×10^{-1}	–2.0	0.4	–/–	8.0×10^{-7}
rs7359257	*IQCH*	Intronic	15	65,489,961	0.45	A/C	0.82	3.9×10^{-9}	4.7×10^{-8}	14,303	6.0×10^{-1}	1.7	0.4	+/–	1.9×10^{-6}

UTR, untranslated region.
[a]Minor allele frequency. [b]Minor/major allele. [c]*P* value for effect heterogeneity between studies. [d]*P* value from stage 1 meta-analysis with genomic control applied to individual studies (up to 87,802 women from 32 studies). [e]*P* value from stage 1 meta-analysis with additional adjustment for overall genomic control. [f]*P* value from *in silico* replication studies (up to 14,731 women). [g]Per allele change in age at menarche (weeks) obtained from a meta-analysis of stage 1 and replication cohorts. [h]Direction of minor allele association with age at menarche in stage 1/replication cohorts. [i]*P* value from meta-analysis of stage 1 (second genomic-control–corrected estimates) and replication cohorts. [j]rs314276 was used as a proxy in the ALSPAC replication sample. [k]These loci reached genome-wide significance in stage 1 but not in the final analysis with second genomic-control correction and combination with replication cohorts.

Table 2 Associations between known obesity-related SNPs and age at menarche

Nearby gene	SNP[a]	Chr.	Obesity phenotype	Menarche β (weeks per allele)	Menarche s.e.	Menarche *P*	Obesity-susceptibility allele	Menarche-decreasing allele
FTO	rs9939609	16q12	BMI	2.5	0.4	3.3×10^{-11}	A	A
SEC16B	rs10913469	1q25	BMI	2.6	0.5	2.4×10^{-8}	C	C
GNPDA2	rs10938397	4p13	BMI	2.1	0.4	8.7×10^{-8}	G	G
NEGR1	rs2815752	1p31	BMI	1.9	0.4	5.9×10^{-7}	A	A
TMEM18	rs6548238	2p25	BMI	2.7	0.5	7.1×10^{-7}	C	C
FAIM2	rs7138803	12q13	BMI	1.8	0.4	1.7×10^{-6}	A	A
BDNF	rs4923461	11p14	BMI	1.7	0.5	3.1×10^{-4}	A	A
KCTD15	rs11084753	19q13	BMI	1.4	0.4	5.9×10^{-4}	G	G
TRA2B, ETV5	rs7647305	3q27	BMI	1.2	0.5	9.0×10^{-3}	C	C
TFAP2B	rs987237	6p12	WHR	1.6	0.5	7.8×10^{-4}	G	G
MSRA	rs7826222	8p23	WHR	1.8	0.8	2.4×10^{-2}	G	G

BMI, body mass index; WHR, waist-hip ratio.
[a]Selected SNPs at each locus are those published for association with BMI, WHR or obesity (rather than those with the strongest signal for age at menarche). SNPs listed are those with a significant association ($P < 0.05$) with age at menarche. A full version of this table including SNPs related to adiposity traits but not reaching significance for menarche can be found in **Supplementary Table 13**.

approach in meta-analysis gene-set enrichment of variant associations (MAGENTA), in which each gene in the genome is assigned an adjusted score that represents its association with age at menarche, and predefined pathways are tested for enrichment of multiple associations (Online Methods). The most significant pathway ($P = 4.9 \times 10^{-3}$) was the biosynthesis of coenzyme A, which is a carrier of acyl groups and is necessary for pyruvate oxidation and fatty acid synthesis and oxidation (**Supplementary Table 9**).

Functional SNP and structural assessment

We explored the potentially functional impacts of our new menarche loci in order to identify their likely genetic mechanisms. In addition, by particularly focusing on those groups of SNPs that have been identified as functional, we aimed to identify possible further menarche loci which did not reach genome-wide significance in our primary meta-analysis.

Copy number variation. Using data from a recent genomic map of copy number variation (CNV)[25], we established that none of the 42 known, confirmed or possible new menarche loci were related to CNVs. Next, we explored the 1,052 CNV-tagging SNPs for association with age at menarche in our GWAS sample. Only one tag SNP was associated with age at menarche after Bonferroni correction (rs3101336, $P = 3 \times 10^{-7}$; **Supplementary Fig. 7**). This SNP tags a CNV near the NEGR1 gene locus, which has been previously associated with body mass index[20].

Non-synonymous SNPs. None of the 42 known, confirmed or possible new menarche variants were amino acid changing. However, two were in strong LD ($r^2 \geq 0.8$) with non-synonymous variants. rs1862471 (intronic in OLFM2 at 19p13.2) is in LD ($r^2 = 0.8$) with rs2303100, which encodes an arginine to glutamine residue change in OLFM2. Second, rs4929923 (in the 3 untranslated region of TRIM66 at 11p15.4) is in LD ($r^2 = 0.92$) with rs11042023, which encodes a histidine to arginine residue change in TRIM66.

To identify possible further menarche loci, we then explored the set of 12,062 non-synonymous SNPs for association with age at menarche in our GWAS sample. Outside of the already associated regions, three non-synonymous SNPs were associated with age at menarche after correction for multiple testing (the Bonferroni threshold for 12,062 independent tests was $P < 4.1 \times 10^{-6}$). These non-synonymous SNPs were rs1254319 in *C14orf39* ($P = 1.9 \times 10^{-7}$), rs7653652 in *C3orf38* ($P = 1.4 \times 10^{-6}$) and rs913588 in *JMJD2C* ($P = 3.3 \times 10^{-6}$).

Expression QTLs. Three of the forty-two known, confirmed or possible new menarche variants were highly significantly cis associated with mRNA expression ($P < 1 \times 10^{-6}$ for mRNA transcript abundance) based on publicly available data from lymphoblastoid cell lines of 400 children (mRNA by SNP Browser). These transcripts were in GAB2 (associated with rs10899489), RBM6 (rs6762477) and NARG2 (rs3743266) (**Supplementary Table 10**). As these genomic loci included a number of genes (**Supplementary Fig. 3**), these specific transcript associations inform the likely functional gene at each locus.

Table 3 Associations between known height SNPs and age at menarche

Gene	SNP[a]	Chr.	Position	Menarche β (weeks per allele)	Menarche s.e.	Menarche *P*	Height-increasing allele	Menarche-increasing allele
LIN28B	rs314277	6	105,514,355	6.9	0.6	2.1×10^{-35}	A	A
PXMP3	rs7846385	8	78,322,734	2.5	0.4	1.9×10^{-9}	C	T
C6orf173	rs4549631	6	127,008,001	1.8	0.4	4.9×10^{-7}	C	T
SCMH1	rs6686842	1	41,303,458	–1.1	0.4	3.3×10^{-3}	T	C
Histone cluster 1	rs10946808	6	26,341,366	1.1	0.4	6.4×10^{-3}	A	A
NOG	rs4794665	17	52,205,328	–0.9	0.4	1.1×10^{-2}	A	G
HMGA2	rs1042725	12	64,644,614	–0.8	0.4	2.0×10^{-2}	C	C
TBX2	rs757608	17	56,852,059	–0.9	0.4	2.2×10^{-2}	A	G
HLA Class III	rs2844479	6	31,680,935	–0.9	0.4	2.4×10^{-2}	A	C
ZBTB38	rs6440003	3	142,576,899	0.8	0.4	3.5×10^{-2}	A	A
CABLES1	rs4800148	18	18,978,326	–1.0	0.5	3.7×10^{-2}	A	G

$\chi^2 = 7.02$, $P = 0.008$ for 11 out of 44 height-associated SNPs also associated with age at menarche (at $P < 0.05$) compared to the 2.2 expected by chance. However, seven height-increasing SNPs are associated with earlier menarche and four are associated with later menarche. Menarche *P* values are derived from our stage 1 meta-analysis of 32 studies with genomic control applied to individual studies.
[a]Selected SNPs at each locus are those published for association with height (rather than those with the strongest signal for age at menarche). SNPs listed are those with a significant association ($P < 0.05$) with age at menarche. A full version of this table including SNPs associated with adult height but not reaching significance for menarche can be found in **Supplementary Table 14**.

Given the likely close biological interaction between the regulation of age at menarche and adiposity, we hypothesized that adipose tissue expressed SNPs (eSNPs) might show a preponderance of associations with age at menarche. Of the 5,184 adipose eSNPs identified in the Icelandic Family Adipose cohort[26], 23 were significantly associated with age at menarche after correction for multiple testing (using a $1/n$ P value threshold for 5,184 independent tests ($P < 1.9 \times 10^{-4}$)) (**Supplementary Table 11**). Of these adipose eSNPs, rs10835211 (menarche $P = 9.4 \times 10^{-6}$) is near *BDNF*, which is a BMI locus and is implicated in eating behavior and body weight regulation[27,28]. rs7160413 (menarche $P = 2.2 \times 10^{-5}$) is near *DLK1*, a gene implicated in early onset puberty[29]. rs133934508 (menarche $P = 3.6 \times 10^{-5}$) is associated with expression of *PITX1*, which encodes a pituitary transcriptional regulator[30].

Candidate gene assessment

Candidate gene studies for age at menarche have largely focused on genes involved in sex steroid-hormone biosynthesis and metabolism, highlighted through animal models or human cases with extreme delayed puberty or hypogonadotrophic hypogonadism[31]. We examined 8,770 SNPs in 16 candidate genes[31–33] and their surrounding regions (±300 kb) for association with age at menarche in our GWAS meta-analysis sample (**Supplementary Table 12**). SNPs in the regions of *TAC3R* (top hit, rs17034046, $P = 3.4 \times 10^{-7}$, ~19 kb upstream of *TAC3R*) and *ESR1* (top hit, rs9383922, $P = 2.2 \times 10^{-6}$, 110 kb upstream of *ESR1*) were significantly associated with age at menarche after correction for multiple testing (the Bonferroni threshold for 8,770 independent tests was $P < 5.7 \times 10^{-6}$). Rare deleterious mutations in *TAC3R*, encoding a receptor for Neurokinin B, and in its ligand *TAC3* have been found in families affected by hypogonadotropic hypogonadism and pubertal failure[31]. *ESR1* encodes an estrogen receptor that is essential for sexual development and reproductive function, and polymorphisms in *ESR1* have previously been nominally associated with age at menarche[33].

Overlapping heritability of body size and menarche timing

Family studies have suggested a substantial coinheritance of the timing of puberty and BMI[34], and this is supported by our finding of four established BMI variants among our new menarche loci. We therefore systematically assessed whether established loci for adiposity-related traits (BMI, waist-hip ratio (WHR) and obesity) and adult height were also associated with age at menarche. Nine of the twelve BMI loci and two of the four WHR loci tested were associated with age at menarche (**Table 2** and **Supplementary Table 13**). In all cases, the BMI- or WHR-increasing allele was associated with earlier menarche, which is consistent with the direction of association in epidemiological studies[35]. Eleven of the forty-four adult height loci were associated with age at menarche (**Table 3** and **Supplementary Table 14**). However, for seven of these loci, the adult height-increasing allele was associated with earlier menarche, which is in the opposite direction to the association in individual-level epidemiological studies[35].

We then assessed the relevance of our new menarche loci to adult BMI and height by exploring *in silico* data from the GIANT consortium. Nine of the forty-two menarche loci were associated with adult BMI (at $P < 0.05$; $N = 32,530$); in all cases, the allele associated with higher BMI was associated with earlier menarche (**Supplementary Table 15**). Eighteen of the menarche loci were associated with adult height (at $P < 0.05$; $N \sim 130,000$); although for three of these loci, the direction of effect was opposite to that predicted from epidemiological studies (**Supplementary Table 16**). Despite these joint associations with body size, in Avon Longitudinal Study of Parents and Children (ALSPAC) mothers, the combined influence of the menarche loci on age at menarche appeared to be completely unattenuated following adjustment for adult height and BMI (**Supplementary Table 17**), suggesting that in general, these menarche loci have direct effects on age at menarche. However, we acknowledge that further large studies with childhood growth data are needed to establish the causal directions of effect of these loci.

DISCUSSION

In a large GWAS meta-analysis comprising over 87,000 women, we identified 30 new loci for the timing of menarche and provide evidence for a further ten possible new loci. These loci were in or near genes associated with cellular development, body weight regulation, hormonal regulation and a wide variety of other biological functions. Previous studies comprising up to 17,510 women had detected only one or two genome-wide significant signals[8–11]. We now show that those earlier signals at *LIN28B* and 9q31.2 represented the 'low-hanging fruit' with particularly large effect sizes relative to their MAF (**Supplementary Fig. 5**). The list of functions of those genes nearest to the menarche loci (**Supplementary Table 7**) and the results of pathway analyses indicate a wide diversity of biological processes that regulate the timing of female pubertal maturation.

Among the confirmed new menarche loci were several loci implicated in body weight regulation, including four loci with established associations with BMI (in or near *FTO*, *SEC16B*, *TRA2B* and *TMEM18*). Furthermore, our systematic analysis of established BMI-related SNPs showed that the majority of alleles related to higher BMI and WHR also showed at least nominal associations with earlier menarche (**Table 2**). It is noteworthy that three new menarche loci are in or near genes implicated in energy homeostasis in animal models (*BSX*, *CRTC1* and *MCHR2*). In the GIANT consortium data, we did not detect any associations between these loci and adult BMI, however the *BSX* and *MCHR2* loci were nominally associated with adult height. In order to robustly investigate whether menarche loci have pleiotropic effects on growth or whether the association with menarche timing is driven through increased adiposity, measures of body fatness before menarche or even before the onset of puberty would be required but were unavailable in most studies. Further functional studies of these new menarche loci may also help to clarify the biological mechanisms linking these traits. In addition to influencing the timing of pubertal initiation, sufficient adiposity is also required for the maintenance of normal hypothalamic-pituitary-gonadal function through signaling by adipocytokines such as leptin[36]. Our pathway analyses highlighted coenzyme A and fatty acid biosynthesis as biological pathways related to menarche timing. Hypothalamic levels of long-chain fatty acyl coenzyme As have been shown to regulate rodent feeding behavior and glucose homeostasis[37], and genetic variants in this pathway could therefore potentially alter central nutrient sensing.

Earlier age at menarche is related to shorter adult stature in large epidemiological studies[35]. We found that several adult height-increasing alleles were also associated with age at menarche (**Table 3**), but at different loci, these alleles were associated with either earlier or later menarche. These paradoxical associations suggest a complex interplay between growth and pubertal timing. Earlier menarche is associated with taller, rather than shorter, childhood height, and there are likely separate causal effects of rapid linear growth on earlier puberty and of earlier pubertal maturation on earlier growth plate fusion and cessation of growth.

Although our pathway analyses strongly identified potential new biological pathways involved in pubertal timing, we acknowledge that the ability to assign putative functions to these menarche loci is substantially limited by the lack of identification of the causal variant at each locus. Many of the strongest associated SNPs were located

hundreds of kilobases distant to the nearest gene, and some menarche loci contained several plausible genes. Indeed, none of the top signals represented non-synonymous SNPs and only two SNPs were in LD with such variants (in *OLFM2* and *TRIM66*). Use of eQTLs helped to identify the likely causal genes (*GAB2*, *RBM6* and *NARG2*) at three menarche loci that spanned multiple genes. However, much future work will be required to identify the causal variants and implicated genes related to these menarche loci.

Despite the large size of our meta-analysis and the substantial increase in the number of menarche loci, these together explained between 3.6%–6.1% of the variance in age at menarche, equivalent to 7.2%–12.2% of its heritability. The majority of menarche loci had estimated effect sizes of between 2 and 3 weeks per allele. Assuming the presence of many true menarche SNPs with an effect size of 2 weeks per allele, even our large meta-analysis would only have had sufficient power to detect half of those SNPs with a MAF of 50% and only one in ten of those SNPs with MAF of 10% (**Supplementary Fig. 8**).

We corrected for population stratification by applying the genomic control method[38] to each of the individual study results. When we applied a more stringent second correction for the overall genomic control inflation factor across all 32 studies, 10 of the 40 possible new menarche variants fell below genome-wide significance (**Fig. 1** and **Table 1**). However, our subsequent finding of confirmatory evidence ($P < 0.05$) even in our limited replication studies for four of these ten variants (in or near *TRIM66*, *TMEM108*, *TMEM18* and *NFAT5*) suggests that the second correction for genomic control is likely to be overconservative.

Our identification of strong associations with SNPs near the candidate genes *TAC3R* and *ESR1* supports the likely presence of further menarche loci which did not meet the genome-wide significance threshold. Systematic assessment of functional genetic variants identified several further putative menarche loci. rs3101336, which tags a CNV near the BMI locus *NEGR1*, showed strong, but not genome-wide significant, association with age at menarche ($P = 3 \times 10^{-7}$). Exploration of adipose tissue eQTLs also identified further putative menarche loci related to genes implicated in eating behavior (*BDNF*), precocious puberty (*DLK1*) and pituitary function (*PITX1*). It has been suggested that lower levels of statistical significance may be applied to variants with prior biological candidacy, however this must be balanced against the desire to avoid false positives, and we suggest that these putative menarche loci require confirmation in further studies.

Notably all of the top menarche variants had MAF ≥ 7%. Although it has been suggested that low-frequency variants have larger effects than common variants[39], our study was clearly underpowered to detect low-frequency variants (MAF < 5%) with modest effect sizes. It is also possible that rare variants are not well captured using genome-wide chips. Future imputation using deep sequencing data from the 1000 Genomes Project may identify additional low frequency hits as well as refine the location of possible functional variants.

In the majority of studies contributing to this report, age at menarche was recalled several years later and often to the nearest completed whole year. Although recalled age at menarche is a valid measure[40] and is unlikely to show systematic bias by genotype, any nondifferential error would lead to reduced statistical power. Menarche indicates the completion of puberty in females, and it is unclear whether our new menarche loci also influence timing of other pubertal phenotypes. The known menarche locus in *LIN28B* was shown to also influence the onset of breast development in girls, the timing of pubic hair development and voice breaking in boys[9] and the timing of the pubertal growth spurt in both boys and girls[41]. Although our new menarche loci might also regulate such wider pubertal processes, it is plausible that some (for example, *INHBA*) might have sex-specific effects. Our study was restricted to cohorts of European ancestry and our results are therefore not generalized to other groups. African-American girls tend to show earlier pubertal maturation compared to girls of European ancestry[42], and genetic studies in such populations might reveal different menarche loci.

In summary, we identified at least 30 new loci for age at menarche. Our findings demonstrate the role of genes which regulate energy homeostasis and hormone pathways and illustrate the complexity of the regulation of the timing of puberty.

URLs. KBiosciences, http://www.kbioscience.co.uk; MACH, http://www.sph.umich.edu/csg/abecasis/MaCH/; METAL, www.sph.umich.edu/csg/abecasis/metal; mRNA by SNP Browser, http://www.sph.umich.edu/csg/liang/asthma/; MAGENTA, http://www.broadinstitute.org/mpg/magenta/; PANTHER, http://www.pantherdb.org/.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

Academy of Finland (Finnish Centre of Excellence in Complex Disease Genetics 129680, 120315, 129287, 129494); Affymetrix (N02-HL-6-4278); Agency of Science, Technology and Research of Singapore (A*STAR); Althingi (the Icelandic Parliament); American Heart Association (0855082E); Amgen; Augustinus Foundation; Australian National Health and Medical Research Council (241944, 339462, 389927, 389875, 389891, 389892, 389938, 442915, 442981, 496739, 552485, 552498, 619667); Australian Research Council (A7960034, A79801419, A79906588, DP0212016, DP0343921, DP0770096); Baltimore Veterans Administration Medical Center Geriatrics Research; Canadian Institutes of Health Research (grant ID 166067); Cancer Research United Kingdom; the Cariplo Foundation; Center for Disease Control and Prevention (USA), Centre for Medical Systems Biology (CMSB); the Chief Scientist Office of the Scottish Government; Clinical Nutrition Research Unit of Maryland (P30 DK072488); Danish National Research Foundation; Danish Pharmacists' Fund; the Egmont Foundation; Erasmus Medical Center; Erasmus University; Estonian Government (SF0180142s08); the European Commision (212111 BBMRI, 205419 ECOGENE, 201413 ENGAGE, HEALTH-F2-2008-201865 - GEFOS, HEALTH-F2-2008-35627, HEALTH-F4-2007-201413, HEALTH-F4-2007-201550 HYPERGENES, 245536 OPENGENE, TREAT-OA, GenomEUtwin Project QLG2-CT-2002-01254, EU/QLRT-2001-01254; ERC-230374); European Union framework program 6 EUROSPAN project (LSHG-CT-2006-018947, LSHM-CT-2003-503041); European Regional Development Fund; Faculty of Biology and Medicine of Lausanne, Switzerland; Fondazione Compagnia di San Paolo Health Ministry, Framingham Heart Study (N01-HC-25195); GENEVA Coordinating Center (U01HG004446); German Federal Ministry of Education and Research; German National Genome Research Network (NGFN-2 and NGFNPlus: 01GS0823); Giorgi-Cavaglieri Foundation; GlaxoSmithKline; Health Fund of the Danish Health Insurance Societies; Helmholtz Zentrum München-German Research Center for Environmental Health; Hjartavernd (the Icelandic Heart Association); Italian Ministry of Health (ICS110.1/RF97.71, RF-FSR-2007-647201); Juvenile Diabetes Research Foundation International (JDRF); Leenaards Foundation; March of Dimes Birth Defects Foundation; the Medical Research Council (G0000934, G0500539, G0701863); National Cancer Institute (CA40356, CA98233, P01CA087969, P01CA055075, CA047988, CA63464, CA54281, CA136792, CA089392, CA104021); Munich Center of Health Sciences (MC Health, LMUinnovativ); National Health and Medical Research Council of Australia (572613 and 003209); National Heart, Lung, and Blood Institute (HL 043851, HL087679, HL69757, N01-HC-55015, N01-HC-55016, N01-HC-55018, N01-HC-55019, N01-HC-55020, N01-HC-55021, N01-HC-55022, R01HL087641, R01HL59367, R01HL086694, RC2 HL102419); National Human Genome Research Institute (NHGRI); National Institute of Aging (263 MD 9164, 263 MD 821336, N.1-AG-1-1, N.1-AG-1-2111, N01-AG-5-0002, AG-16592, Genetics of Reproductive Life Period and Health Outcomes - R21AG032598); National Institute of Arthritis and Musculoskeletal and Skin Diseases and National Institute on Aging ((NIAMS/NIA) R01 AR/AG 41398); National Institute of Child Health and Human Development (HD-061437); National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK, R01DK058845, U01 DK062418); National Institute on Alcohol Abuse and Alcoholism (NIAAA); National Institute of Allergy and Infectious Diseases (NIAID); National Institute on

Drug Abuse (NIDA); National Institutes of Health (AA07535, AA10248, AA13320, AA13321, AA13326, AA14041, HHSN268200782096C, M01 RR-00750, MH66206, N01-AG-12100, N01-AG-1-2109, P01-AG-18397, R01-088119, R01-DA013423, RFAHG006033, U01DE018993, U01DE018903, U01HG004422, U01HG004446, U01HL72515, U01 HL84756, U01HG004399, U01HG004402, U01HG004415, U01HG004423, U01HG004436, U01HG004438, U01HG004402, U01HG004446, U01HG004726, U01HG004728, U01HG004729, U01HG004735, U01HG004738, U01HG04424, U10AA008401, U54RR025204-01, UL1RR025005, UL1 RR025774, Z01CP010200, HHSN268200625226C); the National Institute for Health Research Cambridge Biomedical Research Centre; Netherlands Organization for Scientific Research (904-61-193, 575-25-006, 480-04-004, 56-464-14192, NWO 480-05-003, 175.010.2005.011, 911-03-012, 014-93-015); Netherlands Genomic Initiative (NGI, 050-060-810), National Institute on Aging (NIA) Intramural Research Program; and Laboratory of Neurogenetics; US National Institutes of Health Genes, Environment and Health Initiative; Republic of Croatia Ministry of Science, Education and Sports (108-1080315-0302); Robert Dawson Evans Endowment; the Royal Society, Royal Swedish Academy of Science; State of Bavaria, Germany; Susan G. Komen Breast Cancer Foundation; Swedish Foundation for Strategic Research (SSF); Swedish Heart-Lung Foundation; Swedish Ministry of Higher Education and Research; Swedish Research Council; Swedish Society of Medicine; Swiss National Science Foundation (Ref: 33CSCO-122661, 3100AO-116323/1); the University of Bristol, University of Maryland General Clinical Research Center (M01 RR 16500); the Wellcome Trust (068545/Z/02, 076467/Z/05/Z, 077016/Z/05/Z, 89061/Z/09/Z, strategic award 079895); Western Australian DNA Bank; Western Australian Genetic Epidemiology Resource (see **Supplementary Note** for extended acknowledgments).

AUTHOR CONTRIBUTIONS

Writing group: D.I.B., H.A.B., D.I.C., D.L.C., L.C., E.W.D., M.J.E., C.E.E., T.E., B.F., N.F., F.G., D.F.G., C. He, T.B.H., D.J.H., D.L.K., K.K.L., M. Mangino, M. Marongiu, J.M.M., A. Metspalu, A. Murray, K.K.O., J.R.B.P., L.S., E.A.S., P.S., U.T., A.G.U., C.M.v.D., S.H.v.W., J.A.V., E.W., G.Z.

Analysis working group: C.E.E., A. Murray, K.K.O., J.R.B.P., P.S.

Secondary analyses: J.A.V., J.R.B.P., A.D.J., D.L., A.S.P., V.E., A.V. Segrè, The GIANT Consortium.

Oversight of contributing cohorts: D.I.B., H.A.B., L.J.B., L.C., C.M.v.D., D.F.E., M.J.E., U.d.F., P.G., A.H., D.J.H., P.H., T.B.H., K.-T.K. M. Marongiu, A. Metspalu, A. Murray, G.W.M., J.M.M., S.S.M., V.M., C.E.P., P.P., I.R., P.M.R., E.A.S., D.P.S., G.D.S., K.S., N.J.S., T.D.S., D.T., A.G.U., A.F.W., D.M.W., E.W., H.E.W., J.F.W., N.J.W.

Analytical lead of contributing studies: E.A., E.M.B., D.I.C., D.L.C., T.C., C.E.E., T.E., B.F., N.F., D.F.G., C. Hayward, C. He, J.-J.H., E.I., D.L.K., Z.K., K.L.L., P.L., R.J.F.L., M. Mangino, P.F.M., P.K.E.M., K.N., E.P., J.R.B.P., A.V. Smith, E.N.S., L.S., P.S., S.S., S.U., J.A.V., N.M.W., S.H.v.W., J.H.Z., L.Z.

Genotyping and phenotyping of contributing studies: H.A., N.A., P. d'Adamo, T.A., G.S.B., H.B., I.B., E.B., F.B., J.E.B., S. Bandinelli, S. Bergmann, A.D.C., D.C., H.C., M.C.C., S.J.C., W.C., A.D., R.M.v.D., P.D., G.E., J.E., A.C.H., S.E.H., A.R.F., C.G., L. Ferreli, L. Ferrucci, T.F., F.G., E.J.C.d.G., M.G., V.G., D.G.H., F.B.H., T.I., M.-R.J., D.K., D.P.K., I.K., P.K., T.O.K., J.L., J.S.E.L., L.J.L., S.L., J.B.J.v.M., J.C.M., J.M.M., M. Melbye, N.G.M., W.L.M., A.R.N., F.N., M.N., M.A.N., P.N., K.K.O., B.A.O., A. Palotie, A. Pouta, A.N.P., C.E.P., G.P., L.P., L.J.P., M.P., N.L.P., O.P., F.R., J.P.R., S.M.R., T.R., A.S., A.R.S., C.S., D.S., N.S., S.R.S., U.S., V.S., E.T., K.T., L.T., M.-L.T., J.T., M.U., P.V., G. Waeber, G. Willemsen, M.N.W., L.Y., G.Z., W.V.Z.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at http://www.nature.com/naturegenetics/.

Published online at http://www.nature.com/naturegenetics/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Parent, A.S. *et al.* The timing of normal puberty and the age limits of sexual precocity: variations around the world, secular trends, and changes after migration. *Endocr. Rev.* **24**, 668–693 (2003).
2. Kvale, G. Reproductive factors in breast cancer epidemiology. *Acta Oncol.* **31**, 187–194 (1992).
3. Purdie, D.M. & Green, A.C. Epidemiology of endometrial cancer. *Best Pract. Res. Clin. Obstet. Gynaecol.* **15**, 341–354 (2001).
4. Ong, K.K. *et al.* Earlier mother's age at menarche predicts rapid infancy growth and childhood obesity. *PLoS Med.* **4**, e132 (2007).
5. He, C. *et al.* Age at menarche and risk of type 2 diabetes: results from 2 large prospective cohort studies. *Am. J. Epidemiol.* **171**, 334–344 (2010).
6. Lakshman, R. *et al.* Early age at menarche associated with cardiovascular disease and mortality. *J. Clin. Endocrinol. Metab.* **94**, 4953–4960 (2009).
7. Towne, B. *et al.* Heritability of age at menarche in girls from the Fels Longitudinal Study. *Am. J. Phys. Anthropol.* **128**, 210–219 (2005).
8. He, C. *et al.* Genome-wide association studies identify loci associated with age at menarche and age at natural menopause. *Nat. Genet.* **41**, 724–728 (2009).
9. Ong, K.K. *et al.* Genetic variation in *LIN28B* is associated with the timing of puberty. *Nat. Genet.* **41**, 729–733 (2009).
10. Perry, J.R. *et al.* Meta-analysis of genome-wide association data identifies two loci influencing age at menarche. *Nat. Genet.* **41**, 648–650 (2009).
11. Sulem, P. *et al.* Genome-wide association study identifies sequence variants on 6q21 associated with age at menarche. *Nat. Genet.* **41**, 734–738 (2009).
12. Raivio, T. & Dunkel, L. Inhibins in childhood and puberty. *Best Pract. Res. Clin. Endocrinol. Metab.* **16**, 43–52 (2002).
13. Crofton, P.M. *et al.* Changes in dimeric inhibin A and B during normal early puberty in boys and girls. *Clin. Endocrinol.* **46**, 109–114 (1997).
14. Sehested, A. *et al.* Serum inhibin A and inhibin B in healthy prepubertal, pubertal, and adolescent girls and adult women: relation to age, stage of puberty, menstrual cycle, follicle-stimulating hormone, luteinizing hormone, and estradiol levels. *J. Clin. Endocrinol. Metab.* **85**, 1634–1640 (2000).
15. Burger, H.G. Evidence for a negative feedback role of inhibin in follicle stimulating hormone regulation in women. *Hum. Reprod.* **8**(Suppl 2), 129–132 (1993).
16. Sulzbacher, S., Schroeder, I.S., Truong, T.T. & Wobus, A.M. Activin A-induced differentiation of embryonic stem cells into endoderm and pancreatic progenitors—the influence of differentiation factors and culture conditions. *Stem Cell Rev.* **5**, 159–173 (2009).
17. Dolle, P. Developmental expression of retinoic acid receptors (RARs). *Nucl. Recept. Signal.* **7**, e006 (2009).
18. Frayling, T.M. *et al.* A common variant in the *FTO* gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* **316**, 889–894 (2007).
19. Thorleifsson, G. *et al.* Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat. Genet.* **41**, 18–24 (2009).
20. Willer, C.J. *et al.* Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat. Genet.* **41**, 25–34 (2009).
21. Sakkou, M. *et al.* A role for brain-specific homeobox factor Bsx in the control of hyperphagia and locomotory behavior. *Cell Metab.* **5**, 450–463 (2007).
22. Altarejos, J.Y. *et al.* The Creb1 coactivator Crtc1 is required for energy balance and fertility. *Nat. Med.* **14**, 1112–1117 (2008).
23. Pissios, P., Bradley, R.L. & Maratos-Flier, E. Expanding the scales: The multiple roles of MCH in regulating energy balance and other biological functions. *Endocr. Rev.* **27**, 606–620 (2006).
24. Wu, M., Dumalska, I., Morozova, E., van den Pol, A. & Alreja, M. Melanin-concentrating hormone directly inhibits GnRH neurons and blocks kisspeptin activation, linking energy balance to reproduction. *Proc. Natl. Acad. Sci. USA* **106**, 17217–17222 (2009).
25. Conrad, D.F. *et al.* Origins and functional impact of copy number variation in the human genome. *Nature* **464**, 704–712 (2009).
26. Emilsson, V. *et al.* Genetics of gene expression and its effect on disease. *Nature* **452**, 423–428 (2008).
27. Kernie, S.G., Liebl, D.J. & Parada, L.F. *BDNF* regulates eating behavior and locomotor activity in mice. *EMBO J.* **19**, 1290–1300 (2000).
28. Xu, B. *et al.* Brain-derived neurotrophic factor regulates energy balance downstream of melanocortin-4 receptor. *Nat. Neurosci.* **6**, 736–742 (2003).
29. Temple, I.K., Shrubb, V., Lever, M., Bullman, H. & Mackay, D.J. Isolated imprinting mutation of the *DLK1/GTL2* locus associated with a clinical presentation of maternal uniparental disomy of chromosome 14. *J. Med. Genet.* **44**, 637–640 (2007).
30. Drouin, J., Lamolet, B., Lamonerie, T., Lanctot, C. & Tremblay, J.J. The PTX family of homeodomain transcription factors during pituitary developments. *Mol. Cell. Endocrinol.* **140**, 31–36 (1998).
31. Topaloglu, A.K. *et al. TAC3* and *TACR3* mutations in familial hypogonadotropic hypogonadism reveal a key role for Neurokinin B in the central control of reproduction. *Nat. Genet.* **41**, 354–358 (2009).
32. Gajdos, Z.K. *et al.* Association studies of common variants in 10 hypogonadotropic hypogonadism genes with age at menarche. *J. Clin. Endocrinol. Metab.* **93**, 4290–4298 (2008).
33. Stavrou, I., Zois, C., Ioannidis, J.P. & Tsatsoulis, A. Association of polymorphisms of the oestrogen receptor alpha gene with the age of menarche. *Hum. Reprod.* **17**, 1101–1105 (2002).
34. Kaprio, J. *et al.* Common genetic influences on BMI and age at menarche. *Hum. Biol.* **67**, 739–753 (1995).
35. Onland-Moret, N.C. *et al.* Age at menarche in relation to adult height: the EPIC study. *Am. J. Epidemiol.* **162**, 623–632 (2005).
36. Welt, C.K. *et al.* Recombinant human leptin in women with hypothalamic amenorrhea. *N. Engl. J. Med.* **351**, 987–997 (2004).
37. Pocai, A. *et al.* Restoration of hypothalamic lipid sensing normalizes energy and glucose homeostasis in overfed rats. *J. Clin. Invest.* **116**, 1081–1091 (2006).
38. Devlin, B., Bacanu, S.A. & Roeder, K. Genomic Control to the extreme. *Nat. Genet.* **36**, 1129–1130 author reply 1131 (2004).
39. Manolio, T.A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).
40. Must, A. *et al.* Recall of early menstrual history and menarcheal body size: after 30 years, how well do women remember? *Am. J. Epidemiol.* **155**, 672–679 (2002).
41. Widén, E. *et al.* Distinct variants at LIN28B influence growth in height from birth to adulthood. *Am. J. Hum. Genet.* **86**, 773–782 (2010).
42. Herman-Giddens, M.E. *et al.* Secondary sexual characteristics and menses in young girls seen in office practice: a study from the Pediatric Research in Office Settings network. *Pediatrics* **99**, 505–512 (1997).

Cathy E Elks[1,106], John R B Perry[2,106], Patrick Sulem[3,106], Daniel I Chasman[4,5], Nora Franceschini[6], Chunyan He[7,8], Kathryn L Lunetta[9,10], Jenny A Visser[11], Enda M Byrne[12,13], Diana L Cousminer[14], Daniel F Gudbjartsson[3], Tõnu Esko[15–17], Bjarke Feenstra[18], Jouke-Jan Hottenga[19], Daniel L Koller[20], Zoltán Kutalik[21,22], Peng Lin[23], Massimo Mangino[24], Mara Marongiu[25], Patrick F McArdle[26], Albert V Smith[27,28], Lisette Stolk[11,29], Sophie H van Wingerden[30], Jing Hua Zhao[1], Eva Albrecht[31], Tanguy Corre[32], Erik Ingelsson[33], Caroline Hayward[34], Patrik K E Magnusson[33], Erin N Smith[35–37], Shelia Ulivi[38], Nicole M Warrington[39], Lina Zgaga[40], Helen Alavere[15], Najaf Amin[30], Thor Aspelund[27,28], Stefania Bandinelli[41], Inês Barroso[42], Gerald S Berenson[43], Sven Bergmann[21,22], Hannah Blackburn[42], Eric Boerwinkle[44], Julie E Buring[4,5,45], Fabio Busonero[25], Harry Campbell[40], Stephen J Chanock[46], Wei Chen[43], Marilyn C Cornelis[47], David Couper[48], Andrea D Coviello[49], Pio d'Adamo[38], Ulf de Faire[50], Eco J C de Geus[19], Panos Deloukas[42], Angela Döring[31], George Davey Smith[51], Douglas F Easton[52,53], Gudny Eiriksdottir[27], Valur Emilsson[54], Johan Eriksson[55–58], Luigi Ferrucci[59], Aaron R Folsom[60], Tatiana Foroud[20], Melissa Garcia[61], Paolo Gasparini[38], Frank Geller[18], Christian Gieger[31], The GIANT Consortium[62], Vilmundur Gudnason[27,28], Per Hall[33], Susan E Hankinson[45,63], Liana Ferreli[25], Andrew C Heath[64], Dena G Hernandez[65], Albert Hofman[29,66], Frank B Hu[45,47,63], Thomas Illig[31], Marjo-Riitta Järvelin[67], Andrew D Johnson[9,68], David Karasik[69], Kay-Tee Khaw[70], Douglas P Kiel[69], Tuomas O Kilpeläinen[1], Ivana Kolcic[71], Peter Kraft[45,47,63], Lenore J Launer[61], Joop S E Laven[72], Shengxu Li[1], Jianjun Liu[73], Daniel Levy[9,68,74], Nicholas G Martin[75], Wendy L McArdle[76], Mads Melbye[18], Vincent Mooser[77], Jeffrey C Murray[78], Sarah S Murray[35–37], Michael A Nalls[79], Pau Navarro[34], Mari Nelis[15–17], Andrew R Ness[80], Kate Northstone[76], Ben A Oostra[30,81], Munro Peacock[82], Lyle J Palmer[39], Aarno Palotie[14,42,83], Guillaume Paré[4,5,84], Alex N Parker[85], Nancy L Pedersen[33], Leena Peltonen[14,42,55,83,105], Craig E Pennell[86], Paul Pharoah[52,53], Ozren Polasek[71,87], Andrew S Plump[88], Anneli Pouta[55], Eleonora Porcu[25], Thorunn Rafnar[3], John P Rice[23], Susan M Ring[76], Fernando Rivadeneira[11,29,66], Igor Rudan[40,89], Cinzia Sala[32], Veikko Salomaa[55], Serena Sanna[25], David Schlessinger[90], Nicholas J Schork[35–37], Angelo Scuteri[25,91], Ayellet V Segrè[83,92], Alan R Shuldiner[26,93], Nicole Soranzo[24,42], Ulla Sovio[67], Sathanur R Srinivasan[43], David P Strachan[94], Mar-Liis Tammesoo[15], Emmi Tikkanen[14,55], Daniela Toniolo[32], Kim Tsui[85], Laufey Tryggvadottir[95], Jonathon Tyrer[52,53], Manuela Uda[25], Rob M van Dam[47,96,97], Joyce B J van Meurs[11,29], Peter Vollenweider[98], Gerard Waeber[98], Nicholas J Wareham[1], Dawn M Waterworth[77], Michael N Weedon[2], H Erich Wichmann[31,99,100], Gonneke Willemsen[19], James F Wilson[40], Alan F Wright[34], Lauren Young[85], Guangju Zhai[24], Wei Vivian Zhuang[10], Laura J Bierut[23], Dorret I Boomsma[19], Heather A Boyd[18], Laura Crisponi[25], Ellen W Demerath[60], Cornelia M van Duijn[29,30], Michael J Econs[20,82], Tamara B Harris[61], David J Hunter[45–47,63], Ruth J F Loos[1], Andres Metspalu[15–17], Grant W Montgomery[101], Paul M Ridker[4,5,45,102], Tim D Spector[24], Elizabeth A Streeten[26], Kari Stefansson[3,103], Unnur Thorsteinsdottir[3,103], André G Uitterlinden[11,29,66], Elisabeth Widen[14], Joanne M Murabito[9,49,107], Ken K Ong[1,104,107] & Anna Murray[2,107]

[1]Medical Research Council (MRC) Epidemiology Unit, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK. [2]Genetics of Complex Traits, Peninsula Medical School, University of Exeter, UK. [3]deCODE Genetics, Reykjavik, Iceland. [4]Division of Preventive Medicine, Brigham and Women's Hospital, Boston, Massachusetts, USA. [5]Harvard Medical School, Boston, Massachusetts, USA. [6]Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA. [7]Department of Public Health, Indiana University School of Medicine, Indiana, USA. [8]Melvin and Bren Simon Cancer Center, Indiana University, Indiana, USA. [9]The National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, Massachusetts, USA. [10]Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, USA. [11]Department of Internal Medicine, Erasmus Medical Center (MC), Rotterdam, The Netherlands. [12]Queensland Statistical Genetics, Queensland Institute of Medical Research, Brisbane, Australia. [13]The University of Queensland, Brisbane, Australia. [14]Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Finland. [15]Estonian Genome Center, University of Tartu, Tartu, Estonia. [16]Department of Biotechnology, Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia. [17]Genotyping Core Facility, Estonian Biocenter, Tartu, Estonia. [18]Department of Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark. [19]Department of Biological Psychology, VU University Amsterdam, Amsterdam, The Netherlands. [20]Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, Indiana, USA. [21]Department of Medical Genetics, University of Lausanne, Lausanne, Switzerland. [22]Swiss Institute of Bioinformatics, Lausanne, Switzerland. [23]Department of Psychiatry, Washington University School of Medicine, St. Louis, Missouri, USA. [24]Department of Twin Research and Genetic Epidemiology, King's College London, London, UK. [25]Istituto di Neurogenetica e Neurofarmacologia, Consiglio Nazionale delle Ricerche, Cagliari, Italy. [26]Division of Endocrinology, Diabetes and Nutrition, University of Maryland School of Medicine, Baltimore, Maryland, USA. [27]Icelandic Heart Association, Kopavogur, Iceland. [28]University of Iceland, Reykjavik, Iceland. [29]Netherlands Consortium of Healthy Aging, Rotterdam, The Netherlands. [30]Genetic-Epidemiology Unit, Department of Epidemiology, Erasmus University Medical Center, Rotterdam, The Netherlands. [31]Institute of Epidemiology, Helmholtz Zentrum München-German Research Center for Environmental Health, Neuherberg, Germany. [32]Division of Genetics and Cell Biology, San Raffaele Scientific Institute, Milan, Italy. [33]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [34]MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, Western General Hospital, Edinburgh, UK. [35]Scripps Genomic Medicine, La Jolla, California, USA. [36]The Scripps Translational Science Institute, La Jolla, California, USA. [37]The Scripps Research Institute, La Jolla, California, USA. [38]Medical Genetics, Department of Reproductive Sciences and Development, University of Trieste, Trieste, Italy. [39]Centre for Genetic Epidemiology and Biostatistics, University of Western Australia, Crawley, Australia. [40]Centre for Population Health Sciences, University of Edinburgh, Teviot Place, Edinburgh, Scotland. [41]Geriatric Unit, Azienda Sanitaria di Firenze, Florence, Italy. [42]Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK. [43]Tulane University, New Orleans, Louisiana, USA. [44]Human Genetics Center, University of Texas Health Science Center at Houston, Houston, Texas, USA. [45]Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA. [46]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, Maryland, USA. [47]Department of Nutrition, Harvard School of Public Health, Boston, Massachusetts,

USA. [48]Collaborative Studies Coordinating Center, Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA. [49]Sections of General Internal Medicine, Preventive Medicine and Endocrinology, Department of Medicine, Boston University School of Medicine, Boston, Massachusetts, USA. [50]Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden. [51]MRC Centre for Causal Analyses in Translational Epidemiology, Department of Social Medicine, University of Bristol, Bristol, UK. [52]Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. [53]Department of Oncology, Strangeways Research Laboratories, University of Cambridge, Cambridge, UK. [54]Molecular Programming and Research Informatics (MPRI), Merck and Co., Inc., Rahway, New Jersey, USA. [55]National Institute for Health and Welfare, Helsinki, Finland. [56]Department of General Practice and Primary Health Care, University of Helsinki, Helsinki, Finland. [57]Helsinki University Central Hospital, Unit of General Practice, Helsinki, Finland. [58]Folkhalsan Research Centre, Helsinki, Finland. [59]Longitudinal Studies Section, Clinical Research Branch, National Institute on Aging, Baltimore, Maryland, USA. [60]Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, Minnesota, USA. [61]Laboratory of Epidemiology, Demography and Biometry, Intramural Research Program, National Institute on Aging, Bethesda, Maryland, USA. [62]A full list of members is provided in the **Supplementary Note**. [63]Channing Laboratory, Department of Medicine, Brigham and Women's Hospital, and Harvard Medical School, Boston, Massachusetts, USA. [64]Department of Psychiatry, Washington University School of Medicine, St. Louis, Missouri, USA. [65]Laboratory of Neurogenetics, National Institute of Aging, Bethesda, Maryland, USA. [66]Department of Epidemiology, Erasmus MC, Rotterdam, The Netherlands. [67]Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK. [68]National Heart, Lung, and Blood Institute (NHLBI) Center for Population Studies, Bethesda, Maryland, USA. [69]Hebrew Senior Life Institute for Aging Research and Harvard Medical School, Boston, Massachusetts, USA. [70]Department of Public Health and Primary Care, Institute of Public Health, University of Cambridge, Cambridge, UK. [71]Medical School, University of Zagreb, Zagreb, Croatia. [72]Department of Obstetrics and Gynaecology, Erasmus, MC, Rotterdam, The Netherlands. [73]Human Genetics, Genome Institute of Singapore, Singapore. [74]Division of Cardiology, Boston University School of Medicine, Boston, Massachusetts, USA. [75]Genetic Epidemiology, Queensland Institute of Medical Research, Brisbane, Australia. [76]Avon Longitudinal Study of Parents and Children (ALSPAC), Department of Social Medicine, University of Bristol, Bristol, UK. [77]Genetics Division, GlaxoSmithKline, King of Prussia, Pennsylvania, USA. [78]Department of Pediatrics, University of Iowa, Iowa City, Iowa, USA. [79]Laboratory of Neurogenetics, Intramural Research Program, National Institute on Aging, Bethesda, Maryland, USA. [80]Department of Oral and Dental Science, University of Bristol, Bristol, UK. [81]Department of Clinical Genetics, Erasmus University Medical Center, Rotterdam, The Netherlands. [82]Department of Medicine, Indiana University School of Medicine, Indianapolis, Indiana, USA. [83]Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA. [84]Genetic and Molecular Epidemiology Laboratory, McMaster University, Hamilton, Ontario, Canada. [85]Amgen, Cambridge, Massachusetts, USA. [86]School of Women's and Infants' Health, The University of Western Australia, Crawley, Australia. [87]Gen-Info Ltd., Zagreb, Croatia. [88]Cardiovascular Disease, Merck Research Laboratory, Rahway, New Jersey, USA. [89]Croatian Centre for Global Health, University of Split Medical School, Split, Croatia. [90]Gerontology Research Center, National Institute on Aging, Baltimore, Maryland, USA. [91]Unità Operativa Geriatria-Istituto Nazionale Ricovero e Cura per Anziani (INRCA), IRCCS, Rome, Italy. [92]Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts, USA. [93]Geriatric Research and Education Clinical Center, Veterans Administration Medical Center, Baltimore, Maryland, USA. [94]Division of Community Health Sciences, St. George's, University of London, London, UK. [95]Icelandic Cancer Registry, Reykjavik, Iceland. [96]Department of Epidemiology and Public Health, Yong Loo Lin School of Medicine, National University of Singapore, Singapore. [97]Department of Medicine, National University of Singapore, Singapore. [98]Department of Internal Medicine, BH-10 Centre Hospitalier Universitaire Vaudois (CHUV), Lausanne, Switzerland. [99]Institute of Medical Informatics, Biometry and Epidemiology, Chair of Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany. [100]Klinikum Grosshadern, Munich, Germany. [101]Molecular Epidemiology, Queensland Institute of Medical Research, Brisbane, Australia. [102]Division of Cardiology, Brigham and Women's Hospital, Boston, Massachusetts, USA. [103]Faculty of Medicine, University of Iceland, Reykjavik, Iceland. [104]Department of Paediatrics, University of Cambridge, Cambridge, UK. [105]Deceased. [106]These authors contributed equally to this work. [107]These senior authors jointly supervised this work. Correspondence should be addressed to A. Murray (anna.murray@pms.ac.uk), K.K.O. (ken.ong@mrc-epid.cam.ac.uk) or J.M.M. (murabito@bu.edu).

ONLINE METHODS

Stage 1 GWAS populations. Thirty-two studies contributed to the stage 1 GWAS meta-analysis, comprising 87,802 women of European ancestry. The consortium was made up of populations from the Age, Gene/Environment Susceptibility Study (AGES, $n = 1849$), the Amish population (Amish, $n = 557$), the Atherosclerosis Risk in Communities study (ARIC, $n = 4247$), the British 1958 Birth Cohort (B58C-T1DGC and B58C-WTCCC, $n = 1584$), CoLaus ($n = 2797$), deCODE ($n = 15,864$), the Danish National Birth Cohort (DNBC, $n = 1748$), the Estonian Genome Center, University of Tartu (EGCUT, $n = 987$), the European Prospective Investigation into Cancer and Nutrition (EPIC-obesity cases and cohort, $n = 1840$), the Erasmus Rucphen Family Study (ERF, $n = 1103$), the Framingham Heart Study (FHS, $n = 3801$), the Helsinki Birth Cohort (HBCS, $n = 976$), the Health 2000 study (Health 2000 cases and controls, $n = 922$), InCHIANTI ($n = 597$), the Indiana University premenopausal Caucasian women peak BMD study (Indiana, $n = 1497$), the Nurse's Heath Studies (NHS, $n = 5360$), the Northern Finland Birth cohort (NFBC, $n = 2648$), the Netherlands Twin Register (NTR, $n = 1051$), the Queensland Institute of Medical Research (QIMR, $n = 3528$), the Rotterdam studies (RS1, RS2 and RS3, $n = 5406$), the Study of Addiction: Genetics and Environment (SAGE, $n = 1376$), the SardiNIA study ($n = 2158$), Twins UK I, II and III ($n = 3962$), and the Women's Genome Health Study (WGHS, $n = 22,028$). Full details can be found in the **Supplementary Note**. All studies were approved by local ethics committees and all participants provided written informed consent.

Phenotype measurement and inclusion criteria. Age at menarche recalled by the participant was recorded in each study. Specific questions asked can be found in **Supplementary Table 1**. Only women of European ancestry with a valid age at menarche between 9 and 17 years were included in this analysis, as this represents the normal physiological range. Information on birth year was also collected in each study.

Genotyping. The 32 stage 1 studies were genotyped using a variety of Affymetrix (6.0, GeneChip 500K, 250K, MIP50K and 10K) and Illumina (HumanHap 550K, 318K, HumanHap 300K, HumanHap 370K CNV, HumanHap610 quad, Human660W-Quad BeadChip, 6K and Human 1Mv1_C) genotyping arrays. Genotyping call rate cutoffs were at least 90%, and SNPs were filtered for those with a minor allele frequency of greater than 1%. More details on the filtering criteria for genotypes in each individual study can be found in **Supplementary Table 2**.

Genotype imputation. 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 35 or 36. Algorithms were used to infer unobserved genotypes in a probabilistic manner in either MACH, IMPUTE[43], or software that was developed by the researchers. As a quality control measure, we excluded non-genotyped SNPs with an imputation quality less than 0.3 (for observed versus expected variance in MACH) or 0.4 (for IMPUTE's proper info statistic) from the meta-analysis.

Association testing. Each study performed genome-wide association testing for age at menarche across approximately 2.5 million SNPs based on linear regression under an additive genetic model. Analyses were adjusted for birth year in order to remove the effect of the temporal decline in age at menarche. Studies used PLINK, ProABEL, MACH2QTL, SNPTEST, R packages or MERLIN-fastassoc for the association testing. The results from individual studies were corrected by their respective genomic inflation factors (λ) (**Supplementary Table 1**) according to the genomic control method to correct for population stratification[38].

Meta-analysis. We used an inverse-variance meta-analysis to test the effects of each genetic variant on age at menarche across the 32 studies. Fixed effects models were used, although in the absence of significant heterogeneity, choice of model has little impact on the results. In order to correct for potential relatedness between two Icelandic cohorts (AGES and deCODE), the corrected association results for these cohorts were first meta-analyzed and the genomic-control method was reapplied to the results of the combined sample. These results were then meta-analyzed with the remaining 30 studies. We also displayed further results following a second correction for genomic control using the overall genomic inflation factor calculated from the meta-analysis of all 32 studies. All meta-analyses were conducted using the METAL software package. We considered P values $< 5 \times 10^{-8}$ to indicate genome-wide significance.

We also meta-analyzed results from X-chromosome SNPs in a subset of studies with this data available. This included seven imputed datasets and one directly genotyped dataset. Total sample size was ~60% of the autosomal meta-analysis ($N = 52,781$) and the same statistical model was tested.

Conditional analysis. In order to establish whether genome-wide significant SNPs with low LD in the same chromosomal region (defined as $r^2 < 0.05$ in a 750-kb region) were independent loci, we carried out a conditional analysis. Each study performed a genome-wide analysis for age at menarche using linear regression adjusting for the top signal at each of the 42 associated regions to determine whether potential second signals remained significant even after adjusting for these variants. Birth year was also included as a covariate. Results from each individual study were meta-analyzed to determine whether these potential second signals were truly independent (that is, if $P < 5.0 \times 10^{-8}$).

Replication studies. In order to confirm our possible new menarche loci, we tested our 42 top hits for *in silico* association with age at menarche in 8,669 women from 16 studies with GWAS data and which were not included in the first stage meta-analysis (**Supplementary Table 4**). In addition, new genotype data was generated for 30 of the 42 menarche loci and tested for association with age at menarche in up to 6,118 women from the Avon Longitudinal Study of Parents and Children (ALSPAC). Genotyping was performed by KBiosciences (Hoddesdon, UK) using their own unique system of fluorescence-based competitive allele-specific PCR (KASPar). As in stage 1, analyses were restricted to women reporting age at menarche between 9 and 17 years and adjustment was made for birth year. Mean age at menarche ranged from 12.4 to 13.5 years, consistent with studies in the stage 1 meta-analysis. Linear regression was used to test the association between each variant and age at menarche in an additive genetic model. These results were then meta-analyzed with genomic control–adjusted statistics from our stage 1 meta-analysis using inverse-variance fixed effects models.

In order to calculate the overall variance explained by these menarche loci in each of the replication cohorts, we calculated the r^2 value from a model including all 42 known, confirmed and possible new menarche variants and birth year and compared this to a model including birth year alone. We only included cohorts with >800 women in their full model analyses, as sample sizes smaller than this may give spurious results.

Pathway analysis. Ingenuity pathway analysis (IPA) Knowledge Base 8.5 (Ingenuity Systems) was used to explore the functional relationship between proteins encoded by the 42 known, confirmed and possible new menarche loci. The IPA Knowledge Base contains millions of findings curated from the literature. Genes or nearest genes to the 42 loci (**Table 1**) were entered into the Ingenuity database. These 'focus genes' were analyzed for direct interactions only. Networks were generated with a maximum size of 35 genes and shown as graphical representations of the molecular relationships between genes and gene products. Proteins are depicted as nodes in various shapes representing the functional class of the protein. The biological relationships between nodes are depicted by lines. To determine the probability of the analyzed gene to be found together in a network from Ingenuity Pathways Knowledge Base due to random chance alone, IPA applies a Fisher's exact test. The network score or P value represents the significance of the focus gene enrichment. There are 25 diseases and disorders categories and 32 molecular and cellular function categories in the IPA Knowledge Base. Enrichment of focus genes to these diseases and functional categories was also evaluated. The P value, based on a right-tailed Fisher's exact test, considers the number of identified focus genes and the total number of molecules known to be associated with these categories in the IPA knowledge database.

MAGENTA was used to explore pathway-based associations in the full GWAS dataset. MAGENTA implements a GSEA-based approach, the methodology of which has been previously described[44]. 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

doi:10.1038/ng.714

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 < 0.05 in either analysis (**Supplementary Table 9**). In total, 2,529 pathways from Gene Ontology, PANTHER, KEGG and Ingenuity were tested for enrichment of multiple modest associations with age at menarche.

eQTLs. We tested the association between 5,184 adipose tissue eSNPs identified in the Icelandic Family Adipose (IFA) cohort ($n = 673$) with age at menarche in our stage 1 meta-analysis sample. The IFA cohort dataset included the expression of 23,720 transcripts representing 84% of the 20,060 protein-coding genes annotated in the Ensembl database (v 33)[26].

43. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.* **39**, 906–913 (2007).
44. Segrè, A.V. *et al*. Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLoS Genet.* **6**, e1001058 (2010).

OPEN ACCESS Freely available online

PLoS one

Evaluation of Association of *HNF1B* Variants with Diverse Cancers: Collaborative Analysis of Data from 19 Genome-Wide Association Studies

Katherine S. Elliott[1], Eleftheria Zeggini[1,2], Mark I. McCarthy[1,3], Julius Gudmundsson[4], Patrick Sulem[4], Simon N. Stacey[4], Steinunn Thorlacius[4], Laufey Amundadottir[5], Henrik Grönberg[6], Jianfeng Xu[7,8], Valerie Gaborieau[9], Rosalind A. Eeles[10], David E. Neal[11], Jenny L. Donovan[12], Freddie C. Hamdy[13], Kenneth Muir[14], Shih-Jen Hwang[15], Margaret R. Spitz[16], Brent Zanke[17,18], Luis Carvajal-Carmona[1], Kevin M. Brown[19], Australian Melanoma Family Study Investigators[¤], Nicholas K. Hayward[20], Stuart Macgregor[20], Ian P. M. Tomlinson[1], Mathieu Lemire[17], Christopher I. Amos[16], Joanne M. Murabito[21], William B. Isaacs[22], Douglas F. Easton[23], Paul Brennan[9], The PanScan Consortium[¤], Rosa B. Barkardottir[24,25], Daniel F. Gudbjartsson[4], Thorunn Rafnar[4], David J. Hunter[26], Stephen J. Chanock[5], Kari Stefansson[4], John P. A. Ioannidis[27,28]*

1 Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom, **2** Applied Statistical Genetics, Wellcome Trust Sanger Institute, University of Cambridge, Cambridge, United Kingdom, **3** Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, United Kingdom, **4** deCODE Genetics, Reykjavik, Iceland, **5** Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland, United States of America, **6** Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden, **7** Center for Cancer Genomics, Wake Forest University School of Medicine, Winston-Salem, North Carolina, United States of America, **8** Center for Human Genomics, Wake Forest University School of Medicine, Winston-Salem, North Carolina, United States of America, **9** Genetic Epidemiology Group, International Agency for Research on Cancer (IARC), Lyon, France, **10** Oncogenetics Team, The Institute of Cancer Research, Sutton, United Kingdom, **11** Department of Oncology, University of Cambridge, Cambridge, United Kingdom, **12** Department of Social Medicine, University of Bristol, Bristol, United Kingdom, **13** Nuffield Department of Surgery, University of Oxford, Oxford, United Kingdom, **14** Health Sciences Research Institute, University of Warwick, Coventry, United Kingdom, **15** Framingham Study, National Heart, Lung, and Blood Institute, Bethesda, Maryland, United States of America, **16** Department of Epidemiology, M. D. Anderson Cancer Center, Houston, Texas, United States of America, **17** Ontario Institute for Cancer Research, MaRS Centre, Toronto, Ontario, Canada, **18** Ottawa Health Research Institute, University of Ottawa, Ottawa, Ontario, Canada, **19** Integrated Cancer Genomics Division, The Translational Genomics Research Institute, Phoenix, Arizona, United States of America, **20** Queensland Institute of Medical Research, Royal Brisbane Hospital, Brisbane, Queensland, Australia, **21** Section of General Internal Medicine, Boston University School of Medicine, Boston, Massachusetts, United States of America, **22** The Brady Urological Institute, Johns Hopkins Medical Institutions, Baltimore, Maryland, United States of America, **23** Cancer Research UK Genetic Epidemiology Unit, University of Cambridge, Cambridge, United Kingdom, **24** Department of Pathology, Landspitali-University Hospital of Iceland, Reykjavik, Iceland, **25** Faculty of Medicine, University of Iceland, Reykjavik, Iceland, **26** Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, United States of America, **27** Department of Hygiene and Epidemiology, University of Ioannina School of Medicine and Biomedical Research Institute, Foundation for Research and Technology-Hellas, Ioannina, Greece, **28** Center for Genetic Epidemiology and Modelling, Tufts University School of Medicine, Boston, Massachusetts, United States of America

Abstract

Background: Genome-wide association studies have found type 2 diabetes-associated variants in the *HNF1B* gene to exhibit reciprocal associations with prostate cancer risk. We aimed to identify whether these variants may have an effect on cancer risk in general versus a specific effect on prostate cancer only.

Methodology/Principal Findings: In a collaborative analysis, we collected data from GWAS of cancer phenotypes for the frequently reported variants of *HNF1B*, rs4430796 and rs7501939, which are in linkage disequilibrium ($r^2 = 0.76$, HapMap CEU). Overall, the analysis included 16 datasets on rs4430796 with 19,640 cancer cases and 21,929 controls; and 21 datasets on rs7501939 with 26,923 cases and 49,085 controls. Malignancies other than prostate cancer included colorectal, breast, lung and pancreatic cancers, and melanoma. Meta-analysis showed large between-dataset heterogeneity that was driven by different effects in prostate cancer and other cancers. The per-T2D-risk-allele odds ratios (95% confidence intervals) for rs4430796 were 0.79 (0.76, 0.83)] per G allele for prostate cancer ($p<10^{-15}$ for both); and 1.03 (0.99, 1.07) for all other cancers. Similarly for rs7501939 the per-T2D-risk-allele odds ratios (95% confidence intervals) were 0.80 (0.77, 0.83) per T allele for prostate cancer ($p<10^{-15}$ for both); and 1.00 (0.97, 1.04) for all other cancers. No malignancy other than prostate cancer had a nominally statistically significant association.

Conclusions/Significance: The examined *HNF1B* variants have a highly specific effect on prostate cancer risk with no apparent association with any of the other studied cancer types.

Citation: Elliott KS, Zeggini E, McCarthy MI, Gudmundsson J, Sulem P, et al. (2010) Evaluation of Association of *HNF1B* Variants with Diverse Cancers: Collaborative Analysis of Data from 19 Genome-Wide Association Studies. PLoS ONE 5(5): e10858. doi:10.1371/journal.pone.0010858

Editor: Florian Kronenberg, Innsbruck Medical University, Austria

Received April 8, 2010; **Accepted** April 28, 2010; **Published** May 28, 2010

Funding: JPAI is supported by funding for the Tufts Clinical and Translational Science Institute and the Center for Genetic Epidemiology and Modeling is supported by grant UL1 RR025752 from the National Institutes of Health. KSE is supported by the Wellcome Trust (WT075491/Z/04). EZ is supported by the Wellcome Trust (WT088885/Z/09/Z). NH and SM were funded by the Australian National Health and Medical Research Council (NHMRC) and the National Institutes of Health (grant R01-CA88363). HG is supported by Swedish Cancer Society (Cancerfonden), Swedish Research Council, and Linneus grant "Predication and prevention of breast cancer and prostate cancer". RE and KM are supported by Cancer Research UK Grant C5047/A8385; C5047/A7357, The Institute of Cancer Research and The Everyman Campaign and The Prostate Cancer Research Foundation. The authors acknowledge support from the NIHR to the Biomedical Research Centre at The Institute of Cancer Research and The Royal Marsden NHS Foundation Trust. DEN, JLD and FCH would like to acknowledge the support of the National Cancer Research Institute (NCRI) formed by the Department of Health, the Medical Research Council and Cancer Research UK. The Prostate Testing for Cancer and Treatment (ProtecT) study is funded by the UK National Institute for Health Research Health Technology Assessment Programme (projects 96/20/06, 96/20/99). The NCRI provided funding through ProMPT (Prostate Mechanisms of Progression and Treatment) and this support is gratefully acknowledged. The Cambridge prostate biorepository also received funding from the NIHR Comprehensive Biomedical Research Centre Grant. This support is gratefully acknowledged. The views and opinions expressed therein are those of the authors and are not necessarily those of the funding bodies. deCODE acknowledges grant support from the 6th Framework Program of the EU: contract 018827 (Polygene). The Australian Melanoma Family Study (AMFS) was supported by grants from the NHMRC, Cancer Councils of Queensland, NSW and Victoria, and by National Institutes of Health grant RO1-CA83115 to the GenoMEL consortium. The Framingham Heart Study (FHS) 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.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: jioannid@cc.uoi.gr

¤ See acknowledgments for author list.

Introduction

A large number of epidemiological studies have suggested correlations between type 2 diabetes (T2D) and various cancers[1,2,3]. Most evidence suggests an inverse correlation between T2D and prostate cancer[4,5,6] although not all studies agree on this[7]. Several studies also suggest positive correlations between other cancers and T2D[1,2,3]. It is unclear whether these correlations, if true, represent casual relationships and whether they may also reflect some shared genetic background. Recently, with the advent of genome-wide association studies (GWAS), a large number of genetic variants have been identified that confer susceptibility to T2D or specific types of cancer[8]. An interesting observation has been that specific variants in the *HNF1B* gene (formerly *TCF2*) have been demonstrated to be associated both with the risk of prostate cancer[9,10,11] and the risk of T2D[9,12] with the effects being in the opposite direction for these two phenotypes.

HNF1B was previously known to be mutated in individuals with maturity-onset diabetes of the young type 5 (MODY 5)[13], but a biological explanation of the impact of the identified common variation on T2D and prostate cancer risk remains elusive. The identified genetic effects are small in magnitude even for prostate cancer and T2D, representing odds ratios [ORs] per allele in the range of 1.2 [9,11] and 0.9 [9,12], respectively. Therefore, small effects for other cancer types would not be readily detectable, unless very large studies were performed or data were combined from several studies.

A definitive answer on whether *HNF1B* variants modulate also the risk of other malignancies, or show specificity for prostate cancer, requires large sample sizes. Here we present the results of a large collaborative meta-analysis of *HNF1B*, rs4430796 and rs7501939, which have the most consistent associations with both prostate cancer and T2D. Relevant data were collected on the two variants from GWAS on cancer phenotypes in Caucasian populations in order to examine whether they have an effect on cancer risk in general, on few specific cancers, or only on prostate cancer.

Results

Database of contributed information

All the originally contacted investigators of cancer-related GWA studies agreed to participate in this collaborative analysis, with the exception of the investigators of 3 GWA studies [14,15,16] (1 on breast cancer, 1 on colorectal cancer and 1 on neuroblastoma), 1 of which had no data on the requested variants, as they had used a Affymetrix platform[15]. Investigators who agreed to participate in the collaborative analysis contributed data on 13 datasets for rs4430796 and 19 datasets for rs7501939 [11,17,18,19, 20,21,22,23,24,25,26,27,28,29,30,31,32,33]. For 5 datasets, data were available only for the latter polymorphism either because the polymorphism was not available on the platform used or the SNP failed quality control criteria.

The contributing teams and datasets are shown in Table 1 with data on the number of cases and controls for each polymorphism and for each type of cancer. Datasets from the Framingham cohort contained imputed data for both polymorphisms since an Affymetrix platform had been used, rs4430796 data from the M.D. Anderson Cancer Center was imputed since this SNP had not been directly genotyped, and melanoma data from AMFS and Q-MEGA contained counts from pooling experiments, otherwise all other datasets had direct genotyping on individual participants. Detailed demographic and other characteristics of the study populations can be found in the respective primary publications of these GWA studies [14,15,16,17,18,20,21,22,23,24,25,26,27,28, 29,30,31,32,34].

Overall, the collaborative analysis included data on rs4430796 for 19,640 cancer cases and 21,929 controls; for prostate cancer there were 11,145 cases and 9,650 controls, while for all other cancers there were 8,495 cases and 12,279 controls. The collected data on rs7501939 included 26,923 cases and 49,085 controls; for prostate cancer there were 12,898 cases and 40,371 controls, while for the other cancers there were 14,025 cases and 43,893 controls. Malignancies other than prostate cancer in these datasets included colorectal, breast, lung and pancreatic cancers, and melanoma (Table 1). deCODE contributed data on 4 different cancers and had a common population control group for all 4 of them. The Framingham Heart Study (FHS) contributed data on 4 different cancers and had a common population control group (subjects $\geq$65 years at the last contact who are not nuclear family member of the cancer cases) for all 4 studies with the exception of prostate and breast cancer which used male and female only controls respectively. The common control groups for deCODE and FHS are only counted once in the total sample sizes above.

The meta-analysis of all datasets (Table 2, Figure 1) showed a per T2D risk allele association with both rs4430796 (G allele OR 0.91 [95% CI: 0.88, 0.94] $p = 3\times10^{-10}$) and rs7501939 (T allele

Table 1. Characteristics of datasets included in the collaborative meta-analysis.

Study Centre	Cancer	Genotyping platform(s)	rs4430796 #cases	rs4430796 #controls	rs7501939 #cases	rs7501939 #controls
*ARCTIC	colorectal[23]	Sequenom homogenous MassExtend (in house)	1,079	1,089	1,075	1,087
*AMFS	melanoma[17,24]	Illumina 550K (pooled)	490 [P]	427 [P]	490 [P]	427 [P]
Cambridge	breast [33]	Perlegen	387	363	387	363
*CGEMS	prostate[11,25]	Illumina 550K	4,960	5,021	4,869	4,930
*CAPS	prostate[26]	Sequenom (in house)	2,874	1,708	2,865	1,707
*CORGI	colorectal[27]	Illumina 550K	n/a	n/a	900	908
deCODE	breast[28]	Illumina 300K	n/a	n/a	1,815	30,742
deCODE	colorectal[29]	Illumina 300K	n/a	n/a	988	30,742
deCODE	lung[29,30]	Illumina 300K	n/a	n/a	651	30,742
deCODE	prostate[9,31,32]	Illumina 300K	n/a	n/a	1619	30,742
*FHS	breast[34]	Affymetrix 500K and MIPS 50K combined	182[i]	852 [i]	182 [i]	852 [i]
*FHS	colorectal[34]	Affymetrix 500K and MIPS 50K combined	108 [i]	1,498 [i]	108 [i]	1,498 [i]
*FHS	lung[34]	Affymetrix 500K and MIPS 50K combined	90 [i]	1,498 [i]	90 [i]	1,498 [i]
*FHS	prostate[34]	Affymetrix 500K and MIPS 50K combined	190 [i]	646 [i]	190 [i]	646 [i]
*IARC	lung[20,21]	Illumina 300K	641	2,435	1,797	2,378
*JHH	prostate[26]	Sequenom (in house)	1,512	478	1,521	479
*MDACC	lung[22]	Illumina 317K	1,152 [i]	1,137 [i]	1,152	1,137
*PANSCAN	pancreatic Stage 1[19,48]	Illumina 550K and 610K	1,754	1,796	1,757	1,796
*PANSCAN	pancreatic Stage 2[19,48]	Illumina 550K and 610K	1,748	1,818	1,769	1,841
*Q-MEGA	melanoma[24]	Illumina 550K (pooled)	864[P]	864 [P]	864 [P]	864 [P]
*UKGPCS	prostate[18]	Illumina 550K	1,609	1,797	1,834	1,867

Unless otherwise indicated all data is from direct genotyping. *ARCTIC (Assessment of Risk for Colorectal Tumors in Canada), AMFS (Australian Melanoma Family Study), CGEMS (Cancer Genetics Markers of Susceptibility), CAPS (Cancer of the Prostate in Sweden), CORGI (Colorectal Tumour Gene Identification), FHS (Framingham Heart Study), IARC (International Agency for Research on Cancer), JHH (Johns Hopkins Hospital), MDACC (M.D. Anderson Cancer Center, Texas), PANSCAN (Pancreatic Cancer Cohort Consortium), Q-MEGA (Queensland study of Melanoma: Environment and Genetic Associations), UKGPCS (UK Genetic Prostate Cancer Study). n/a: no available data; *i*: imputed; *p*: pooled.

doi:10.1371/journal.pone.0010858.t001

OR 0.91 [95% CI: 0.88, 0.94] $p = 5v10^{-10}$) according to fixed effects calculations, while by random effects calculations there was nominal significance (OR 0.94 [95% CI: 0.88, 1.00], $p = 0.033$ for rs4430796 and 0.93 [95% CI: 0.86, 1.01], $p = 0.07$ for rs7501939). The reason for this diversity is that there was very large between-study heterogeneity in the effect sizes (I^2 of 82% [95% CI: 73-89%] and 80% [95% CI: 70-86%], respectively, for the two polymorphisms; Q-test p-value <0.001 for both polymorphisms), and this makes the fixed effects calculations less reliable. Results were qualitatively similar when we increased the variance in deCODE, FHS, and IARC estimates to account for the overlapping control group (not shown).

The heterogeneity was largely driven by the diversity in the effect sizes between prostate cancer and all other cancers. A meta-analysis limited to prostate cancer datasets gave consistent associations with both rs4430796 (OR per copy of T2D risk allele (A) 0.79 [95% CI: 0.76, 0.83], $p<10^{-15}$ by fixed effects and 0.79 [95% CI: 0.74, 0.84] $p = 10^{-13}$ by random effects), and rs7501939 (OR per copy of T2D risk allele (T) 0.80 [95% CI: 0.77, 0.83] $p<10^{-15}$ by fixed effects and 0.79 [95% CI: 0.74, 0.85], $p = 2\times10^{-11}$ by random effects) (Table 2). There was some residual between-study heterogeneity even within the prostate cancer datasets (I^2 of 42% [95%CI: 0–79%] and 56% [95% CI: 0–82%], respectively, for the two polymorphisms; Q-test p-value 0.037 and 0.14, respectively), although the heterogeneity pertained only to the exact magnitude of the genetic effects and a nominally statistically significant association was seen in each of the datasets except for the Framingham study where the number of prostate cancer cases was more limited.

Conversely, the results for all other cancers suggested no significant association and results were consistent across studies. The summary OR was 1.03 and 1.00 for the two polymorphisms respectively ($p = 0.14$ and 0.81 by fixed effects) and the 95% CIs excluded ORs deviating more than 7% from the null (OR = 1.00) for rs4430796 and more than 4% from the null for rs7501939 (Table 2). The Q-test p-value was 0.99 and 0.45 for the two polymorphisms respectively and random effects estimates were thus identical to fixed effects estimates.

There was also no convincing evidence for an association between either of the two polymorphisms and any of the other cancers (besides prostate cancer), when each cancer type was evaluated separately. Point estimates were in the opposite direction (odds ratio 1.03–1.05) for pancreatic and lung cancer, but were not nominally statistically significant (Table 2). The difference between the prostate cancer and other cancers' effect estimates was beyond chance ($p<0.05$) for both polymorphisms.

Table 2. Summary of results for association between rs4430796 and rs7501939 and diverse cancer types.

	rs4430796	rs4430796	rs4430796	rs7501939	rs7501939	rs7501939
Cancer type	**Studies (cases, controls)**	**OR (95% CI)**	**I^2 (95% CI)**	**Studies (cases, controls)**	**OR (95% CI)**	**I^2 (95% CI)**
All cancers	16 (19,640, 21,929)*	0.91 (0.88, 0.94)	82 (73, 89)	21 (26,923, 49,085)*	0.92 (0.90, 0.95)	80 (70, 86)
Prostate	5 (11,145, 9,650)	0.79 (0.76, 0.83)	42 (0, 79)	6 (12,898, 40,371)	0.80 (0.77, 0.83)	56 (0, 82)
All Others	11 (8,495, 12,279)*	1.03 (0.99, 1.07)	0 (0, 60)	15 (14,025, 43,893)*	1.00 (0.97, 1.04)	0 (0, 54)
Breast	2 (569, 1,215)	1.00 (0.84, 1.20)	n/a	3 (2,384, 31,957)	0.97 (0.91, 1.04)	0 (0, 90)
Lung	3 (1,883, 5,070)	1.05 (0.98, 1.13)	0 (0, 90)	4 (3,690, 35,755)	1.03 (0.96, 1.10)	0 (0, 85)
Colorectal	2 (1,187, 2,587)	1.01 (0.90, 1.14)	n/a	4 (3,071, 34,235)	1.01 (0.94, 1.08)	0 (0, 85)
Melanoma	2 (1,354, 1,291)	0.98 (0.87, 1.01)	n/a	2 (1,354, 1,291)	1.01 (0.90, 1.13)	n/a
Pancreatic	2 (3,502, 3,614)	1.04 (0.98, 1.11)	n/a	2 (3,526, 3,637)	1.03 (0.97, 1.10)	n/a

OR: odds ratio, CI: confidence interval, n/a: not applicable (heterogeneity I^2 confidence intervals are not calculated when there are only 2 studies). Odds ratios are based on fixed effects calculations. When point estimates or confidence intervals differ by over 1% in random effects calculations, random effects results are mentioned in the text. * the common control groups of deCODE and FHS are counted only once.

doi:10.1371/journal.pone.0010858.t002

Discussion

The current collaborative analysis documents that both rs4430796 and rs7501939 have robust support for association with prostate cancer, while we did not observe any convincing evidence for an association of any of the other cancers examined with either polymorphism. When data from all other cancers, excluding prostate cancer, were combined the summary effects had 95% CIs that excluded even subtle associations. Apart from prostate cancer, when other datasets for each individual cancer type was combined, the 95% CIs consistently excluded associations with modest effects. This would suggest that the effects mediated by these polymorphisms are specific to T2D and prostate cancer and they do not involve any other cancer types.

The *HNF1B* gene encodes a transcription factor and it was initially identified as a MODY gene[13]. Subsequent studies have suggested that mutations in this gene may also be related to renal disease[35] and chromophobe renal cell carcinoma[36]. No GWAS evaluating kidney cancer were included in our analysis, and no kidney cancer GWAS has been published to-date. The expression profile of the gene shows a tissue-specific pattern. It is essential for embryonic survival and is expressed in the gut, kidney, liver, lung, pancreas, prostate, thymus and genital tract [37,38]. It could be speculated that the lack of association with some cancers studied here may be due to the low or absent expression of this gene in those tissues (for example breast cancer). We did not have data on liver cancer, thymoma or genital tract cancer, but data on lung, pancreatic, and colorectal cancer showed no association, with point estimates very near to the null.

The two variants that we assessed are not necessarily the functional culprits. GWA studies typically derive markers of phenotypes that are probably linked with the functional genetic variation[39]. However, identifying the functional variants is difficult. Even if they could be identified, it is unlikely that substantially large genetic effects for other cancers would exist, if the tagging variants have so consistently null effects. Another caveat is that we only examined populations of Caucasian descent. This reduces the heterogeneity that could be due to different LD patterns in populations of different ancestry. However, it would be worthwhile to investigate the associations of the *HNF1B* variants for T2D, prostate cancer, and other cancers, also in non-Caucasian populations. Preliminary data suggest that both of the examined variants had consistent associations with T2D in Caucasian, Asian (Hong Kong), and West African ancestry participants[9], while the association of rs4430796 with prostate cancer risk was found to be even stronger in the Japanese than in Caucasian populations[40]. Moreover, it would be useful to dissect associations with specific disease subsets. Even within the analyzed Caucasian-descent populations, we observed some modest between-study heterogeneity in the strength of the association between the *HNF1B* variants and prostate cancer. This may be due to different associations in different sub-phenotypes. For example, some data suggest that the rs4430796 A allele may primarily increase the risk for early-onset (<50 years) prostate cancer rather than later-onset disease[41].

In conclusion, while the two examined *HNF1B* variants conclusively have pleiotropic effects on both T2D and prostate cancer, the pleiotropy apparently does not extend to other cancer types. Genetic effects may offer a way to dissect comorbidity between specific cancers and metabolic phenotypes. Besides *HNF1B*, other gene loci have started appearing where variants are identified that modulate susceptibility to both T2D and some malignancy, e.g. prostate cancer for the *JAZF1* locus gene [11,42] and melanoma for the *CDKN2A* locus [43], although different, unlinked variants are implicated in the susceptibility to the malignancy and T2D, respectively. The elucidation of correlated pleiotropic effects on diverse phenotypes will require very large studies, given the generally subtle effects involved. Collaborative efforts between multiple teams, as in the current study, may offer a suitable approach to answer such questions.

Methods

Eligible GWA investigations and data

We used the NHGRI catalogue of published GWA studies[44], a comprehensive database of GWA investigations to identify GWA studies on cancer phenotypes published as of May 20, 2008. We also performed additional PubMed searches to identify whether any additional GWA studies on cancer phenotypes had been published until then. We focused on solid cancers, excluding hematologic malignancies. Given that these GWAS did not include any studies on pancreatic cancer (of special interest, given the association with T2D), we also identified GWAS on pancreatic cancer that had not been published by that time, so as to ensure their inclusion.

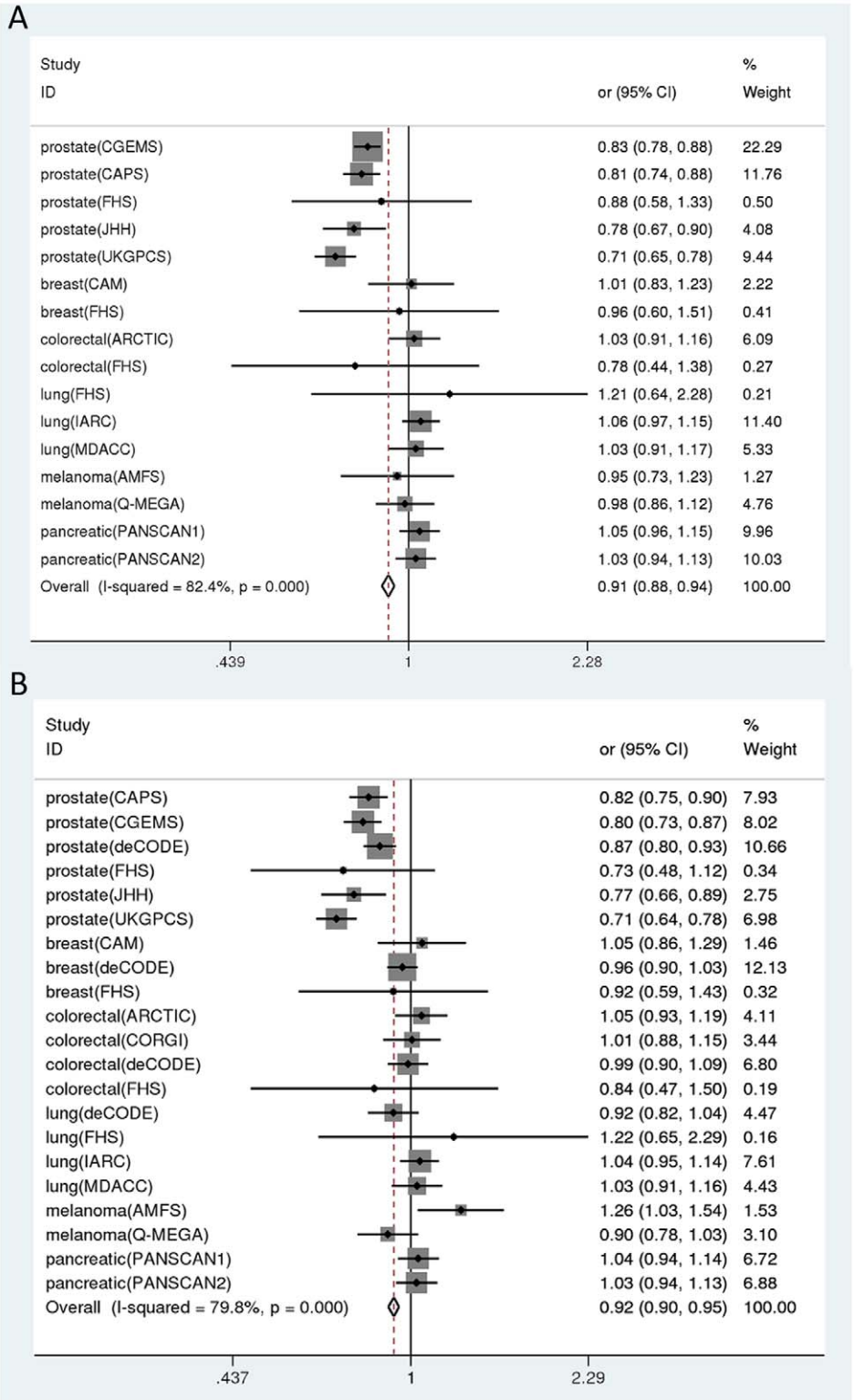


Figure 1. Association of rs4430796 and rs7501939 with diverse cancer types. Panel A shows results for rs4430796 and panel B shows results for rs7501939. Each study is shown by its odds ratio and 95% confidence intervals). Prostate cancer studies appear on the top and other cancer studies follow in alphabetical order. For the abbreviations of the names of the studies see Table 1. The summary diamond at the bottom corresponds to the fixed effects summary. Weight indicates the relative proportion of the total evidence found in each study (the weight is inversely proportional to the variance).

doi:10.1371/journal.pone.0010858.g001

We communicated with the corresponding and principal investigators of all of these studies to request their participation in the collaborative meta-analysis. The investigators of these studies were asked to contribute relevant data on genotype frequencies in cancer cases and non-cancer controls for the *HNF1B* variants, rs4430796 and rs7501939. The risk alleles for prostate cancer are A and C for rs4430796 and rs7501939 respectively. The risk alleles for T2D are G and T for rs4430796 and rs7501939 respectively. The two SNPs have modestly high LD in Caucasians, but low LD in Africans ($r^2 = 0.77$ and 0.22 in CEU and YRI, respectively). Investigators were requested to provide all GWA data that they had obtained for evaluation of any cancer phenotype, including any additional unpublished datasets. Additional genotyping for the two specific variants was encouraged, when a GWA platform had been used that did not directly genotype these polymorphisms (e.g. Affymetrix or Perlegen rather than Illumina). When a study had data on more than one cancer type, data were requested to be provided separately for each cancer type. Investigators were asked to provide also information and clarifications about the design of their studies, and to ensure that population stratification and cryptic relatedness had been appropriately addressed and appropriate quality controls were available for the genotyping. All GWAS investigations that contributed data on these SNPs used stringent QC standards (as described in detail in their original publications) and the two SNPs fulfilled these standards. Approval from local institutional review boards and steering committees was obtained, as deemed necessary for each study by its investigators. The contributed data were checked for completeness and with logical queries and any missing or unclear information was clarified through communication with the contributing investigators.

Meta-analysis

For each SNP, we performed meta-analyses including the data from all eligible cancer studies (regardless of the specific cancer phenotype addressed) and also subgroup meta-analyses, with each subgroup limited to studies addressing a specific cancer phenotype. A separate analysis compared the results of the association for prostate cancer versus the association for all other cancers combined.

All analyses followed the per allele (log-additive) model of inheritance with effect sizes expressed in the odds ratio (OR) scale using both fixed and random effects models[45]. Heterogeneity testing used the Q statistic (considered statistically significant at $p<0.10$), and the I^2 metric[46] and its 95% CIs [47]. Analyses excluding data from studies with pooled genotyping gave similar results (not shown).

Based on the accumulated total sample size and given the minor allele frequencies of these two polymorphisms in HapMap CEU (47% for rs4430796 A allele and 47% for rs7501939 T allele), the meta-analysis had 95% or higher power to detect an association of OR = 1.10 at alpha = 0.05 with each of the two polymorphisms for overall cancer risk, prostate cancer risk, or other cancer risk. Reported p-values are two-tailed. Analyses were performed in STATA 10.0 (College Station, Texas).

Acknowledgments

Australian Melanoma Family Study investigators: Graham J Mann[1], John L Hopper[2], Joanne F Aitken[3], Richard F Kefford[1], Graham G Giles[4], Bruce K Armstrong[5].

[1]Westmead Institute of Cancer Research, University of Sydney at Westmead Millennium Institute and Melanoma Institute Australia, Westmead, New South Wales, Australia. [2]Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, School of Population Health, University of Melbourne, Melbourne, Victoria, Australia. [3]Viertel Centre for Research in Cancer Control, The Cancer Council Queensland, Brisbane, Queensland, Australia. [4]Cancer Epidemiology Centre, The Cancer Council Victoria, Carlton, Victoria, Australia. [5]School of Public Health, University of Sydney, New South Wales, Australia.

PanScan Consortium: Gloria M. Petersen*[1], Laufey Amundadottir*[2,3], Charles S. Fuchs*[4,5], Peter Kraft*[6,7], Rachael Z. Stolzenberg-Solomon*[3], Kevin B. Jacobs[3,8,9], Alan A. Arslan[10,11,12], H. Bas Bueno-de-Mesquita[13], Steven Gallinger[14], Myron Gross[15], Kathy Helzlsouer[16], Elizabeth A. Holly[17], Eric J. Jacobs[18], Alison P. Klein[19,20], Andrea LaCroix[21], Donghui Li[22], Margaret T. Mandelson[21,23], Sara H. Olson[24], Harvey A. Risch[25], Wei Zheng[26], Demetrius Albanes[3], William R. Bamlet[1], Christine D. Berg[27], Marie-Christine Boutron-Ruault[28], Julie E. Buring[29,30], Paige M. Bracci[17], Federico Canzian[31], Sandra Clipp[32], Michelle Cotterchio[33], Mariza de Andrade[1], Eric J. Duell[34], J. Michael Gaziano[35,36], Edward L. Giovannucci[5,6,37], Michael Goggins[38], Göran Hallmans[39], Susan E. Hankinson[5,6], Manal Hassan[22], Barbara Howard[40], David J. Hunter[5,6], Amy Hutchinson[3,8], Mazda Jenab[34], Rudolf Kaaks[31], Charles Kooperberg[21], Vittorio Krogh[41], Robert C. Kurtz[42], Shannon M. Lynch[3], Robert R. McWilliams[1], Julie B. Mendelsohn[3], Dominique S. Michaud[6,43], Hemang Parikh[2,3], Alpa V. Patel[18], Petra H.M. Peeters[43,44], Aleksandar Rajkovic[45], Elio Riboli[43], Laudina Rodriguez[46], Daniela Seminara[47], Xiao-Ou Shu[26], Gilles Thomas[3,48], Anne Tjønneland[49], Geoffrey S. Tobias[3], Dimitrios Trichopoulos[6,50], Stephen K. Van Den Eeden[51], Jarmo Virtamo[52], Jean Wactawski-Wende[53], Zhaoming Wang[3,8], Brian M. Wolpin[4,5], Herbert Yu[25], Kai Yu[3], Anne Zeleniuch-Jacquotte[11,12], Joseph F. Fraumeni, Jr.[3], Robert N. Hoover*[3], Patricia Hartge*[3], Stephen J. Chanock*[2,3].

[1]Department of Health Sciences Research, College of Medicine, Mayo Clinic, Rochester, MN, USA. [2]Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, MD, USA. [3]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department Health and Human Services, Bethesda, MD, USA. [4]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA. [5]Channing Laboratory, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. [6]Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA. [7]Department of Biostatistics, Harvard School of Public Health, Boston, MA, USA. [8]Core Genotyping Facility, SAIC-Frederick Inc., NCI-Frederick, Frederick, MD, USA. [9]Bioinformed Consulting Services, Gaithersburg, MD, USA. [10]Department of Obstetrics and Gynecology, New York University School of Medicine, New York, NY, USA. [11]Department of Environmental Medicine, New York University School of Medicine, New York, NY, USA. [12]New York University Cancer Institute, New York, NY, USA. [13]National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands and Department of Gastroenterology and Hepatology, University Medical Centre Utrecht, Utrecht, The Netherlands. [14]Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada. [15]Department of Laboratory Medicine/Pathology, School of Medicine, University of Minnesota, Minneapolis, MN, USA. [16]Prevention and Research Center, Mercy Medical Center, Baltimore, MD, USA. [17]Department of Epidemiology & Biostatistics, University of California San Francisco, San Francisco, CA, USA. [18]Department of Epidemiology, American Cancer Society, Atlanta, GA, USA. [19]Department of Oncology, the Johns Hopkins University School of Medicine, Baltimore, MD, USA. [20]Department of Epidemiology, the Bloomberg School of Public Health, The Sol Goldman Pancreatic Research Center, The Johns Hopkins Medical Institutions, Baltimore, MD, USA. [21]Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA. [22]Department of Gastrointestinal Medical Oncology, University of Texas M.D. Anderson Cancer Center, Houston, Texas, USA. [23]Group Health Center for Health Studies, Seattle, WA, USA. [24]Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, NY, USA. [25]Yale University School of Public Health, New Haven, CT, USA. [26]Department of Medicine and Vanderbilt-Ingram Cancer Center, Vanderbilt University, Nashville, TN, USA. [27]Division of Cancer Prevention, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA. [28]Inserm, Paris-Sud University, Institut Gustave-Roussy, Villejuif, France. [29]Divisions of Preventive Medicine and Aging, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston,

MA, USA. [30]Department of Ambulatory Care and Prevention, Harvard Medical School, Boston, MA, USA. [31]German Cancer Research Center (DKFZ), Heidelberg, Germany. [32]Johns Hopkins Bloomberg School of Public Health, George W. Comstock Center for Public Health Research and Prevention, Hagerstown, MD, USA. [33]Cancer Care Ontario and Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada. [34]Catalan Institute of Oncology (ICO), Barcelona, Spain. [35]Physicians' Health Study, Divisions of Aging, Cardiovascular Disease, and Preventive Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA. [36]Massachusetts Veterans Epidemiology Research and Information Center, Veterans Affairs Boston Healthcare System, Boston, MA, USA. [37]Department of Nutrition, Harvard School of Public Health, Boston, MA, USA. [38]Departments of Oncology, Pathology and Medicine, The Johns Hopkins University School of Medicine, Baltimore, MD, USA. [39]Department of Public Health and Clinical Medicine, Nutritional Research, Umeå University, Umeå, Sweden. [40]MedStar Research Institute, Georgetown University, Hyattsville, MD [41]Nutritional Epidemiology Unit, Fondazione IRCCS Istituto Nazionale dei Tumoridi Milano, Milan, Italy. [42]Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, NY, USA. [43]Divisionof Epidemiology, Public Health and Primary Care, Imperial College London, London, UK. [44]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands. [45]Departments of Obstetrics and Gynecology and Reproductive Sciences, University of Pittsburgh, Pittsburgh, PA, USA [46]Public Health and Participation Directorate, Health and Health Care Services Council, Asturias, Spain. [47]Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, MD, USA. [48]Synergie-Lyon-Cancer, INSERM U590, Centre Leon Berard, 69373 Lyon, Cedex [08], France [49]Institute of Cancer Epidemiology, Danish Cancer Society, Copenhagen, Denmark. [50]Bureau of Epidemiologic Research, Academy of Athens, Greece [51]Division of Research, Kaiser Permanente, Northern California Region, Oakland, CA, USA. [52]Department of Chronic Disease Prevention, National Institute for Health and Welfare, Helsinki, Finland. [53]Department of Social and Preventive Medicine, University at Buffalo, SUNY Buffalo, NY, USA.

Author Contributions

Conceived and designed the experiments: KSE EZ MIM JPAI. Analyzed the data: KSE JPAI. Contributed reagents/materials/analysis tools: KSE JG PS SNS ST LTA HG JX VG RAE DEN JLD FCH KM SJH MRS BZ LCC KMB AMFSI NKH SM IPT ML CIA JMM WBI DFE PB PC RB DFG TR DJH SJC KS. Wrote the paper: KSE EZ MIM JG JPAI.

References

1. Inoue M, Iwasaki M, Otani T, Sasazuki S, Noda M, et al. (2006) Diabetes mellitus and the risk of cancer: results from a large-scale population-based cohort study in Japan. Arch Intern Med 166: 1871–1877.
2. Stattin P, Bjor O, Ferrari P, Lukanova A, Lenner P, et al. (2007) Prospective study of hyperglycemia and cancer risk. Diabetes Care 30: 561–567.
3. Rousseau MC, Parent ME, Pollak MN, Siemiatycki J (2006) Diabetes mellitus and cancer risk in a population-based case-control study among men from Montreal, Canada. Int J Cancer 118: 2105–2109.
4. Kasper JS, Giovannucci E (2006) A meta-analysis of diabetes mellitus and the risk of prostate cancer. Cancer Epidemiol Biomarkers Prev 15: 2056–2062.
5. Gong Z, Neuhouser ML, Goodman PJ, Albanes D, Chi C, et al. (2006) Obesity, diabetes, and risk of prostate cancer: results from the prostate cancer prevention trial. Cancer Epidemiol Biomarkers Prev 15: 1977–1983.
6. Calton BA, Chang SC, Wright ME, Kipnis V, Lawson K, et al. (2007) History of diabetes mellitus and subsequent prostate cancer risk in the NIH-AARP Diet and Health Study. Cancer Causes Control 18: 493–503.
7. Will JC, Vinicor F, Calle EE (1999) Is diabetes mellitus associated with prostate cancer incidence and survival? Epidemiology 10: 313–318.
8. Manolio TA, Brooks LD, Collins FS (2008) A HapMap harvest of insights into the genetics of common disease. J Clin Invest 118: 1590–1605.
9. Gudmundsson J, Sulem P, Steinthorsdottir V, Bergthorsson JT, Thorleifsson G, et al. (2007) Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. Nat Genet 39: 977–983.
10. Sun J, Zheng SL, Wiklund F, Isaacs SD, Purcell LD, et al. (2008) Evidence for two independent prostate cancer risk-associated loci in the HNF1B gene at 17q12. Nat Genet 40: 1153–1155.
11. Thomas G, Jacobs KB, Yeager M, Kraft P, Wacholder S, et al. (2008) Multiple loci identified in a genome-wide association study of prostate cancer. Nat Genet.
12. Winckler W, Weedon MN, Graham RR, McCarroll SA, Purcell S, et al. (2007) Evaluation of common variants in the six known maturity-onset diabetes of the young (MODY) genes for association with type 2 diabetes. Diabetes 56: 685–693.
13. Horikawa Y, Iwasaki N, Hara M, Furuta H, Hinokio Y, et al. (1997) Mutation in hepatocyte nuclear factor-1 beta gene (TCF2) associated with MODY. Nat Genet 17: 384–385.
14. Tenesa A, Farrington SM, Prendergast JG, Porteous ME, Walker M, et al. (2008) Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. Nat Genet 40: 631–637.
15. Gold B, Kirchhoff T, Stefanov S, Lautenberger J, Viale A, et al. (2008) Genome-wide association study provides evidence for a breast cancer risk locus at 6q22.33. Proc Natl Acad Sci U S A 105: 4340–4345.
16. Maris JM, Mosse YP, Bradfield JP, Hou C, Monni S, et al. (2008) Chromosome 6p22 locus associated with clinically aggressive neuroblastoma. N Engl J Med 358: 2585–2593.
17. Cust AE, Schmid H, Maskiell JA, Jetann J, Ferguson M, et al. (2009) Population-based, case-control-family design to investigate genetic and environmental influences on melanoma risk: Australian Melanoma Family Study. Am J Epidemiol 170: 1541–1554.
18. Eeles RA, Kote-Jarai Z, Giles GG, Olama AA, Guy M, et al. (2008) Multiple newly identified loci associated with prostate cancer susceptibility. Nat Genet 40: 316–321.
19. Petersen GM, Amundadottir L, Fuchs CS, Kraft P, Stolzenberg-Solomon RZ, et al. (2010) A genome-wide association study identifies pancreatic cancer susceptibility loci on chromosomes 13q22.1, 1q32.1 and 5p15.33. Nat Genet 42: 224–228.
20. Brennan P, McKay J, Moore L, Zaridze D, Mukeria A, et al. (2007) Uncommon CHEK2 mis-sense variant and reduced risk of tobacco-related cancers: case control study. Hum Mol Genet 16: 1794–1801.
21. Hung RJ, McKay JD, Gaborieau V, Boffetta P, Hashibe M, et al. (2008) A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. Nature 452: 633–637.
22. Amos CI, Wu X, Broderick P, Gorlov IP, Gu J, et al. (2008) Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. Nat Genet 40: 616–622.
23. Zanke BW, Greenwood CM, Rangrej J, Kustra R, Tenesa A, et al. (2007) Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. Nat Genet 39: 989–994.
24. Brown KM, Macgregor S, Montgomery GW, Craig DW, Zhao ZZ, et al. (2008) Common sequence variants on 20q11.22 confer melanoma susceptibility. Nat Genet 40: 838–840.
25. Yeager M, Orr N, Hayes RB, Jacobs KB, Kraft P, et al. (2007) Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. Nat Genet 39: 645–649.
26. Duggan D, Zheng SL, Knowlton M, Benitez D, Dimitrov L, et al. (2007) Two genome-wide association studies of aggressive prostate cancer implicate putative prostate tumor suppressor gene DAB2IP. J Natl Cancer Inst 99: 1836–1844.
27. Tomlinson IP, Webb E, Carvajal-Carmona L, Broderick P, Howarth K, et al. (2008) A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. Nat Genet 40: 623–630.
28. Stacey SN, Manolescu A, Sulem P, Thorlacius S, Gudjonsson SA, et al. (2008) Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. Nat Genet 40: 703–706.
29. Rafnar T, Sulem P, Stacey SN, Geller F, Gudmundsson J, et al. (2009) Sequence variants at the TERT-CLPTM1L locus associate with many cancer types. Nat Genet 41: 221–227.
30. Thorgeirsson TE, Geller F, Sulem P, Rafnar T, Wiste A, et al. (2008) A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. Nature 452: 638–642.
31. Gudmundsson J, Sulem P, Manolescu A, Amundadottir LT, Gudbjartsson D, et al. (2007) Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. Nat Genet 39: 631–637.
32. Gudmundsson J, Sulem P, Rafnar T, Bergthorsson JT, Manolescu A, et al. (2008) Common sequence variants on 2p15 and Xp11.22 confer susceptibility to prostate cancer. Nat Genet.
33. Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, et al. (2007) Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 447: 1087–1093.
34. http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000007.v9.p4.
35. Edghill EL, Bingham C, Ellard S, Hattersley AT (2006) Mutations in hepatocyte nuclear factor-1beta and their related phenotypes. J Med Genet 43: 84–90.
36. Rebouissou S, Vasiliu V, Thomas C, Bellanne-Chantelot C, Bui H, et al. (2005) Germline hepatocyte nuclear factor 1alpha and 1beta mutations in renal cell carcinomas. Hum Mol Genet 14: 603–614.
37. Reber M, Cereghini S (2001) Variant hepatocyte nuclear factor 1 expression in the mouse genital tract. Mech Dev 100: 75–78.

38. Coffinier C, Barra J, Babinet C, Yaniv M (1999) Expression of the vHNF1/HNF1beta homeoprotein gene during mouse organogenesis. Mech Dev 89: 211–213.
39. Ioannidis JP, Thomas G, Daly MJ (2009) Validating, augmenting and refining genome-wide association signals. Nat Rev Genet 10: 318–329.
40. Yamada H, Penney KL, Takahashi H, Katoh T, Yamano Y, et al. (2009) Replication of prostate cancer risk loci in a Japanese case-control association study. J Natl Cancer Inst 101: 1330–1336.
41. Levin AM, Machiela MJ, Zuhlke KA, Ray AM, Cooney KA, et al. (2008) Chromosome 17q12 variants contribute to risk of early-onset prostate cancer. Cancer Res 68: 6492–6495.
42. Zeggini E, Scott LJ, Saxena R, Voight BF, Marchini JL, 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.
43. Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, et al. (2007) Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. Science 316: 1336–1341.
44. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, et al. (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci U S A 106: 9362–9367.
45. Kavvoura FK, Ioannidis JP (2008) Methods for meta-analysis in genetic association studies: a review of their potential and pitfalls. Hum Genet 123: 1–14.
46. Higgins JP, Thompson SG, Deeks JJ, Altman DG (2003) Measuring inconsistency in meta-analyses. Bmj 327: 557–560.
47. Ioannidis JP, Patsopoulos NA, Evangelou E (2007) Uncertainty in heterogeneity estimates in meta-analyses. Bmj 335: 914–916.
48. Petersen GM, Amundadottir L, Fuchs CS, Kraft P, Stolzenberg-Solomon RZ, et al. A genome-wide association study identifies pancreatic cancer susceptibility loci on chromosomes 13q22.1, 1q32.1 and 5p15.33. Nat Genet.

Periaortic Fat Deposition Is Associated With Peripheral Arterial Disease

The Framingham Heart Study

Caroline S. Fox, MD, MPH; Joseph M. Massaro, PhD; Christopher L. Schlett, MD; Sam J. Lehman, MBBS; James B. Meigs, MD, MPH; Christopher J. O'Donnell, MD, MPH; Udo Hoffmann, MD, MPH; Joanne M. Murabito, MD, ScM

Background—Central obesity is associated with peripheral arterial disease, suggesting that ectopic fat depots may be associated with localized diseases of the aorta and lower-extremity arteries. We hypothesized that persons with greater amounts of periaortic fat are more likely to have clinical PAD and a low ankle-brachial index.

Methods and Results—We quantified periaortic fat surrounding the thoracic aorta using a novel volumetric quantitative approach in 1205 participants from the Framingham Heart Study Offspring cohort (mean age, 65.9 years; women, 54%); visceral abdominal fat also was measured. Clinical peripheral arterial disease was defined as a history of intermittent claudication, and ankle-brachial index was dichotomized as low (≤0.9) or lower-extremity revascularization versus normal (>0.9 to <1.4). Regression models were created to examine the association between periaortic fat and intermittent claudication or low ankle-brachial index (n=66). In multivariable logistic regression, per 1 SD increase in periaortic fat, the odds ratio for the combined end point was 1.52 (P=0.004); these results were strengthened with additional adjustment for body mass index (odds ratio, 1.69; P=0.002) or visceral abdominal fat (odds ratio, 1.67; P=0.009), whereas no association was observed for visceral abdominal fat (P=0.16). Similarly, per SD increase in body mass index or waist circumference, no association was observed after accounting for visceral abdominal fat (body mass index, P=0.35; waist circumference, P=0.49).

Conclusions—Periaortic fat is associated with low ABI and intermittent claudication. **(*Circ Cardiovasc Imaging*. 2010; 3:515-519.)**

Key Words: obesity ■ atherosclerosis ■ peripheral vascular diseases

Peripheral arterial disease (PAD) affects >12% of adults in the United States and is strongly associated with multiple cardiovascular disease risk factors.[1,2] PAD is associated with an increased risk of cardiovascular disease and all-cause mortality,[3,4] highlighting the need for a better understanding of the pathogenesis of PAD.

Clinical Perspective on p 519

Although traditional cardiovascular disease risk factors such as smoking and diabetes also are strong risk factors for PAD,[5] only central obesity, but not generalized obesity, has been shown to be associated with PAD.[6,7] In this context, body fat distribution is an important factor in determining overall cardiometabolic risk.[8,9] Ectopic fat depots, defined as fat depots in nonclassical locations,[10] typically are believed to exert systemic effects on cardiometabolic risk. However, locally acting ectopic fat depots may contribute to obesity-mediated vascular disease.[10,11] In particular, perivascular fat, or fat that surrounds blood vessels, is a physiological modulator of vascular tone and adipocyte hypertrophy that can lead to hypoxia, inflammation, and oxidative stress.[12] Further, recent experimental work suggests that perivascular fat may provide a mechanistic link between metabolic signals and vessel wall inflammation[13] and vascular smooth muscle cell proliferation.[14]

We have developed a reproducible protocol to quantify periaortic fat[15] in order to examine whether perivascular fat may mediate diseases of the aorta. Because prior findings demonstrated an association of central but not generalized obesity with PAD, we hypothesized that persons with greater amounts of periaortic fat will have a higher prevalence of low ankle-brachial index (ABI) values and clinical PAD.

Received August 17, 2009; accepted June 29, 2010.

From the National Heart, Lung, and Blood Institute Framingham Heart Study (C.S.F., C.J.O., J.M. Murabito), Bethesda, Md, and Framingham, Mass; Departments of Cardiac MR PET CT (C.L.S., S.J.L, U.H.) and General Internal Medicine (J.M. Murabito), Massachusetts General Hospital and Harvard Medical School, Boston Mass; Department of Cardiology (S.J.L.), Flinders University, Adelaide, Australia; Department of Mathematics (J.M. Massaro), Boston University, Boston, Mass; Division of Endocrinology (C.S.F.), Brigham and Women's Hospital and Harvard Medical School, Boston, Mass; and Department of Medicine (J.M. Murabito), Section of General Internal Medicine, Boston University School of Medicine, Boston, Mass.

Guest Editor for this article was Joao A. Lima, MD.

Correspondence to Caroline S. Fox, MD, MPH, 73 Mt Wayte Ave, Ste 2, Framingham, MA 01702. E-mail foxca@nhlbi.nih.gov

Circ Cardiovasc Imaging **is available at http://circimaging.ahajournals.org** **DOI: 10.1161/CIRCIMAGING.110.958884**

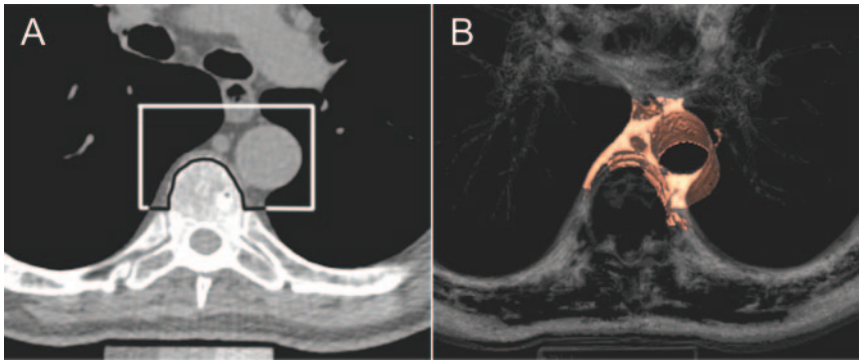


Figure. A, The upper boundaries of periaortic fat in an axial CT image. B, The corresponding 3D reconstruction. Periaortic fat, as measured by CT, was defined as any pixel of attenuation between −195 and −45 Hounsfield units within the region of interest.

Methods

Study Sample

In 1971, the Framingham Offspring Study enrolled children and spouses of the original Framingham Heart Study cohort. Participants for the current analysis took part in the multidetector CT substudy. From the Framingham Offspring Study, 1422 participants underwent chest and abdominal multidetector CT from 2002 to 2005. Of these, 1397 were analyzed for perivascular fat, and 1295 had nonmissing ABI ≤1.4. Of these, 1205 had nonmissing covariates and were included in the analysis.

The institutional review boards of Boston University Medical Center and Massachusetts General Hospital approved the study protocol. Written informed consent was provided by all participants.

Multidetector CT Scan Protocol

Scans of the abdomen and chest were performed with 8-slice multidetector CT. In the chest cavity, a series of 2.5-mm slices were acquired from the level of the carina to the diaphragm during an inspiratory breath-hold using prospective ECG triggering (120 kVp, 320 mA). In the abdomen, 2.5-mm slices (120 kVp, 320 mA) were obtained from the upper edge of the S1 vertebrae and 125 mm superiorly.

Measurement of Periaortic Fat Volume

Image analyses were performed on a dedicated workstation.[15] Using a semiautomated method, adipose tissue quantification was performed, which required manually defining the tissue borders. To calculate adipose tissue volumes, CT attenuation thresholds (window width, −195 to −45 Hounsfield units; window center, −120 Hounsfield units) were used. The anatomic borders to define thoracic periaortic fat were (1) anterior (indicating the area immediately surrounding the thoracic aorta), which was defined by a line drawn horizontally through the esophagus that connected to the left costovertebral joint, and (2) posterior, which was defined by the right lateral border of the vertebral body and the anterior edge of the vertebral body. These definitions resulted in a 6.75-cm column of fat (27 slices) surrounding the thoracic aorta. The Figure shows the region that was quantified. We also defined a measure consisting of abdominal periaortic fat, which consisted of tracing 5-mm rings calibrated to the vessel diameter. However, technical limitations, including the inherent relationship with the vessel diameter and the inability to visualize the retroperitoneal lining, limit the interpretation of these data, and hence, they are not included in the current analysis. Reproducibility was excellent for intra- and interreader measurements of the thoracic periaortic fat (intraclass correlation coefficient, 0.99 and 0.98, respectively).[15]

We quantified visceral abdominal fat (VAT) as previously described.[17] Briefly, the reader manually traced the abdominal muscular wall separating the subcutaneous from the VAT depot. Semiautomatic quantification of fat volumes was facilitated with a window width of −195 to −45 Hounsfield units.

ABI and PAD Assessment

At examination 8 (2005 to 2008), ankle and brachial blood pressures were routinely measured on all participants. Participants rested for a minimum of 5 minutes in the supine position on the examining table before blood pressure measurement. Blood pressure cuffs were applied to bare ankles with the midpoint of the bladder over the posterior tibial artery approximately 3 cm above the medial malleolus. Systolic blood pressure was obtained using a 9.6-MHz Doppler pen probe and an ultrasonic Doppler flow detector. For each limb, the cuff was inflated rapidly to the maximal inflation level and deflated at a rate of 2 mm Hg per second until the systolic blood pressure became audible. Measurements were obtained in the following order: right arm, right ankle, left ankle, left arm. All limb blood pressures were repeated in reverse order. Measurement was obtained from the dorsalis pedis artery only if the posterior tibial pulse could not be located by palpation or with the Doppler pen probe.

The ABI was calculated for each leg as the ratio of the average systolic blood pressure in the ankle divided by the average systolic blood pressure in the arm. The higher arm mean was used to calculate the ABI for each leg. The lower of the ABIs from the 2 legs was used for analysis.

As part of routine Framingham Heart Study research examinations, a physician-administered medical history interview was conducted that included queries about lower-extremity revascularization. Medical records were obtained to verify self-report of all revascularization procedures. The physician also used a standardized questionnaire to query the participant about symptoms of intermittent claudication (IC).[2] IC was defined as exertional discomfort in the calf that appeared sooner with uphill or more rapid walking pace and was relieved with rest. An end point review panel of 3 senior investigators made the final determination of the presence of IC. The mean time between CT scan acquisition and ABI measurement was 2.5 years.

Risk Factor Assessment

Cardiometabolic risk factors also were quantified during the examination 8. Body mass index (BMI) was defined as the weight (kg) divided by height (m^2). Waist circumference (WC) was ascertained at the umbilicus. Fasting morning blood samples were collected for blood glucose and lipids. Diabetes was defined as a fasting plasma glucose of at least 126 mg/dL or hypoglycemic treatment. Current-smokers were defined as participants who smoked at least 1 cigarette per day in the year prior to their eighth examination. The definition of hypertension used was systolic blood pressure of at least 140 mm Hg, diastolic blood pressure of at least 90 mm Hg, or antihypertensive therapy.

Circulating fasting plasma levels of resistin and adiponectin were quantified by ELISA. Intraassay coefficients of variation were 9.0% for resistin and 5.8% for adiponectin.

Statistical Analysis

Thoracic periaortic fat and VAT volumes were normally distributed. ABI was dichotomized at ≤0.9; participants with a history of

Table 1. Study Sample Characteristics Among Participants (n=1205)

Characteristic	Value
Age, y	65.9±8.9
Women	647 (53.7)
BMI, kg/m^2	28.4±5.3
WC, cm	99.1±13.5
Triglycerides, mg/dL*	102 (73–144)
Total/HDL cholesterol, mg/dL	3.52±1.05
Lipid treatment	518 (42.9)
Hypertension	688 (57.1)
Diabetes mellitus	97 (8.1)
Current-smoker	106 (8.8)
Former-smoker	647 (53.7)
Adiponectin, μg/mL†	9.9±6.0
Resistin, ng/mL†	14.6±8.2
Thoracic periaortic fat, cm^3	16.3±9.1
VAT, cm^3	2089.9±1099.9
ABI ≤0.9	45 (3.8)
IC	35 (2.9)
Lower-extremity revascularization‡	7 (0.6)

Data are presented as mean±SD or no. (%), unless otherwise indicated. HDL indicates high-density lipoprotein.

*Median with 25th to 75th percentiles.

†In a subsample of 975 participants.

‡Seven participants with lower-extremity revascularization were part of the 66 participants with IC or low ABI; 3 also had prevalent IC, 1 had an ABI ≤0.9, 1 had both IC and abnormal ABI, and 2 had neither.

lower-extremity surgery were considered in the low ABI category. Participants with an ABI >1.4 were excluded because none had any IC symptoms or a prior revascularization procedure. Low ABI and IC were combined as the primary analysis (n=66). Analyses were modeled with periaortic fat and VAT as exposures, and low ABI and IC were combined as outcomes. BMI and WC also were modeled as separate exposures; all adipose tissue data are presented per 1 SD increase. The multivariable logistic regression model included the covariates of age, sex, smoking, diabetes, hypertension, total and high-density lipoprotein cholesterol, lipid treatment, and log triglycerides. Additional models with periaortic fat included BMI or VAT as covariates. Analyses were performed with SAS version 9.1.3. $P<0.05$ was considered statistically significant.

Results

Study Sample Characteristics

The mean age of the study sample was 65.8 years, and 53.7% were women (Table 1). Overall, 45 participants had an ABI ≤0.9, 35 had a history of IC, and 66 had either a low ABI or IC. Thoracic periaortic fat was strongly associated with VAT ($r=0.74$; $P<0.001$).

Table 3. Multivariable-Adjusted Regressions for Periaortic Fat and Low ABI and IC Individually

	Low ABI (n=45)		IC (n=35)	
Model	OR (95% CI)	*P*	OR (95% CI)	*P*
Age, sex	1.89 (1.42–2.52)	<0.001	1.90 (1.39–2.59)	<0.001
Age, sex, MV	1.78 (1.27–2.48)	<0.001	1.54 (1.08–2.19)	0.02
Age, sex, MV+BMI	2.07 (1.41–3.04)	<0.001	1.62 (1.09–2.41)	0.02
Age, sex, MV+VAT	1.98 (1.25–3.13)	0.004	1.69 (1.05–2.72)	0.03

Data are presented as per 1 SD increase of periaortic fat. The multivariable was adjusted for smoking, diabetes, hypertension, total and high-density lipoprotein cholesterol, lipid treatment, and log triglycerides. MV indicates multivariable.

Association Between Periaortic Fat and Combined Low ABI and IC

In minimally adjusted models per SD increase in periaortic fat, the odds ratio (OR) for low ABI or IC was 1.79 (95% CI, 1.40 to 2.30; $P<0.001$) (Table 2). Further adjustment for clinical covariates modestly affected the OR (1.52; $P=0.004$). Similarly, additional adjustment for BMI or VAT did not materially affect the results (BMI-adjusted OR, 1.69; VAT-adjusted OR, 1.67). In contrast, VAT was associated with low ABI or IC in minimally adjusted models (OR, 1.44; 95% CI, 1.12 to 1.87; $P=0.005$), but these findings were attenuated after adjustment for standard covariates (OR, 1.23; $P=0.16$).

We also examined the associations between BMI and WC with low ABI or IC (Table 2). BMI was not associated with the combined end point in minimally adjusted models (OR, 1.19; $P=0.18$), whereas WC was modestly associated with low ABI or IC in age- and sex-adjusted models (OR, 1.37; $P=0.02$), which was attenuated on adjustment for VAT (OR, 1.14; $P=0.49$).

Association Between Periaortic Fat and Low ABI or IC

Results for low ABI or IC as separate outcomes were similar to the combined outcome models (Table 3).

Secondary Analysis

In secondary analyses of a subset of participants (n=975), we also adjusted our primary model of VAT as a correlate for

Table 2. Multivariable-Adjusted Regressions for Periaortic Fat and the Combined End Point of IC and Low ABI (n=66)

	Thoracic Periaortic Fat		VAT		BMI		WC	
	OR (95% CI)	*P*	OR (95% CI)	*P*	OR (95% CI)	*P*	OR (95% CI)	*P*
Age, sex	1.79 (1.40–2.30)	<0.001	1.44 (1.12–1.87)	0.005	1.19 (0.92–1.52)	0.18	1.37 (1.06–1.77)	0.02
Age, sex, MV	1.52 (1.15–2.02)	0.004	1.23 (0.92–1.65)	0.16	1.03 (0.78–1.36)	0.85	1.23 (0.93–1.65)	0.15
Age, sex, MV+BMI	1.69 (1.22–2.34)	0.002	1.40 (0.95–2.06)	0.09	…	…	1.64 (1.03–2.62)	0.04
Age, sex, MV+VAT	1.67 (1.14–2.45)	0.009	…	…	0.83 (0.57–1.22)	0.35	1.14 (0.78–1.67)	0.49

Data are presented as per 1 SD increase of thoracic abdominal fat, VAT, BMI, or WC. The multivariable was adjusted for smoking, diabetes, hypertension, total and high-density lipoprotein cholesterol, lipid treatment, and log triglycerides. MV indicates multivariable.

low ABI or IC for adiponectin and resistin. The results were not materially different (OR, 1.73; $P=0.01$).

Discussion

Principal Findings

In our community-based sample of participants from the Framingham Heart Study, we demonstrated that periaortic fat is associated with low ABI and IC. We did not observe a similar association with BMI, WC, or VAT. Our findings suggest a potential role for periaortic fat in the pathogenesis of PAD.

In the Context of Current Literature

The association between BMI and PAD has been inconsistent.[18,19] Some studies demonstrate a linear association between BMI and ABI level,[20] whereas others showed no association[5] or an association with the highest BMI in participants within the lowest ABI category[21] or in those with a high ABI (>1.3).[18] Central obesity, but not BMI, has previously been associated with PAD in a cohort of elderly men.[6] Similarly, in a study of elderly participants from the Osteoporotic Fractures in Men study, waist-to-hip ratio, but not BMI, was associated with low ABI.[7] In the German cohort of the Reduction of Atherothrombosis for Continued Health registry, 50% of patients with PAD had abdominal obesity.[22] Obesity previously has been associated with the severity of PAD.[23] Obese patients report more calf pain than the general population, and obese patients who undergo surgical treatment for obesity have a lower risk of developing calf pain.[24] Taken together, the literature suggests that body composition, particularly for persons with increased central fat, may indicate increased risk for PAD. The present study extends these findings by identifying an association between periaortic fat and PAD.

Potential Mechanisms

Experiments in rat aorta demonstrate that periaortic adipose tissue releases growth factors that stimulate smooth muscle cell proliferation that is enhanced in aged rats and rats fed a high-fat diet.[14] These findings suggest that perivascular adipose tissue may promote vascular disease through dysfunction of smooth muscle cells. Adipocytes secrete numerous other factors, including proinflammatory cytokines and adipokines, that also may promote development of vascular disease.[13] Recent in vitro work demonstrated that under basal conditions, human perivascular adipocytes show evidence of a proinflammatory state and reduced adipocyte differentiation. Thus, perivascular adipocytes may contribute to adventitial inflammation and, in turn, the development of atherosclerosis. Greenstein et al[12] isolated perivascular adipose tissue from small arteries taken from gluteal fat biopsy samples and demonstrated that the adipocytes secrete adiponectin, a physiological modulator of vascular tone. On further examination of perivascular fat from obese subjects, the investigators noted a loss of this vasodilatory effect due to adipocyte hypertrophy, leading to inflammation and oxidative stress. In the obese Zucker rat, hind limb blood flow was reduced with concomitant stiffer vessels, and this was independent of muscle mass and physical activity,[25] providing a potential mechanism for which obesity can lead to PAD. The pathophysiologic mechanisms by which local adipose tissue influences development of vascular disease remain to be determined and represent an exciting area of active research.

Clinical and Research Implications

These findings highlight the potential toxic role of periaortic fat on the peripheral vasculature and suggest a potential mechanism whereby obesity might lead to the development of PAD. Further research is necessary to uncover the specific mechanisms of disease. Whether reduction of periaortic fat can lead to reduced PAD or PAD progression requires further examination.

Strengths and Limitations

The strength of the present analysis was the detailed characterization of ectopic fat depots, which allowed us to examine the associations between thoracic periaortic fat and VAT with low ABI and IC. Important covariates were routinely collected, limiting any potential for recall bias. Some limitations warrant mention. First, the participants were white, limiting generalizability to other races/ethnicities. Second, our measure of abdominal periaortic fat is not reliable, limiting our ability to directly quantify this fat depot in the abdomen. We used thoracic periaortic fat as a proxy measure of perivascular fat through the entire arterial tree because we were unable to quantify perifemoral artery fat. Further research is necessary to better understand the distribution of periaortic fat through the vascular territory. Ectopic fat depots are hypothesized to primarily have systemic effects, such as VAT,[8] or local effects, such as pericardial fat or periaortic fat. The results from the present study suggest that only periaortic fat and not BMI, WC, or VAT are associated with PAD, rendering a systemic effect of periaortic fat unlikely. However, this study is cross-sectional; therefore, causation cannot be inferred. Smoking was defined as cigarette smoking within the past 12 months; thus, some degree of misclassification may occur among participants who stopped smoking within this time interval.

Sources of Funding

This work was supported by the National Heart, Lung, and Blood Institute Framingham Heart Study (N01-HC-25195), the General Clinical Research Centers Program (grant number M01-RR-01066), and a Career Development Award from the American Diabetes Association and National Institute of Diabetes and Digestive and Kidney Diseases (K24 DK080140) to Dr Massaro.

Disclosures

Dr Massaro currently has research grants from GlaxoSmithKline and Sanofi-Aventis, and has consulting agreements with Eli Lilly and Interleukin Genetics. The other authors report no conflicts.

References

1. Ostchega Y, Paulose-Ram R, Dillon CF, Gu Q, Hughes JP. Prevalence of peripheral arterial disease and risk factors in persons aged 60 and older: data from the National Health and Nutrition Examination Survey 1999–2004. *J Am Geriatr Soc*. 2007;55:583–589.
2. Murabito JM, D'Agostino RB, Silbershatz H, Wilson WF. Intermittent claudication. A risk profile from The Framingham Heart Study. *Circulation*. 1997;96:44–49.
3. Saw J, Bhatt DL, Moliterno DJ, Brener SJ, Steinhubl SR, Lincoff AM, Tcheng JE, Harrington RA, Simoons M, Hu T, Sheikh MA, Kereiakes DJ,

Topol EJ. The influence of peripheral arterial disease on outcomes: a pooled analysis of mortality in eight large randomized percutaneous coronary intervention trials. *J Am Coll Cardiol*. 2006;48:1567–1572.
4. 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, 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, McDermott MM. Ankle brachial index combined with Framingham Risk Score to predict cardiovascular events and mortality: a meta-analysis. *JAMA*. 2008;300:197–208.
5. Selvin E, Erlinger TP. Prevalence of and risk factors for peripheral arterial disease in the United States: results from the National Health and Nutrition Examination Survey, 1999–2000. *Circulation*. 2004;110:738–743.
6. Planas A, Clara A, Pou JM, Vidal-Barraquer F, Gasol A, de MA, Contreras C, Marrugat J. Relationship of obesity distribution and peripheral arterial occlusive disease in elderly men. *Int J Obes Relat Metab Disord*. 2001;25:1068–1070.
7. Vogt MT, Cauley JA, Kuller LH, Hulley SB. Prevalence and correlates of lower extremity arterial disease in elderly women. *Am J Epidemiol*. 1993;137:559–568.
8. Fox CS, Massaro JM, Hoffmann U, Pou KM, Maurovich-Horvat P, Liu CY, Vasan RS, Murabito JM, Meigs JB, Cupples LA, D'Agostino RB Sr, O'Donnell CJ. Abdominal visceral and subcutaneous adipose tissue compartments: association with metabolic risk factors in the Framingham Heart Study. *Circulation*. 2007;116:39–48.
9. Rosito GA, Massaro JM, Hoffmann U, Ruberg FL, Mahabadi AA, Vasan RS, O'Donnell CJ, Fox CS. Pericardial fat, visceral abdominal fat, cardiovascular disease risk factors, and vascular calcification in a community-based sample: the Framingham Heart Study. *Circulation*. 2008;117:605–613.
10. Montani JP, Carroll JF, Dwyer TM, Antic V, Yang Z, Dulloo AG Ectopic fat storage in heart, blood vessels and kidneys in the pathogenesis of cardiovascular diseases. *Int J Obes Relat Metab Disord.* 2004;28(suppl 4):S58–S65.
11. Yudkin JS, Eringa E, Stehouwer CD. "Vasocrine" signalling from perivascular fat: a mechanism linking insulin resistance to vascular disease. *Lancet*. 2005;365:1817–1820.
12. Greenstein AS, Khavandi K, Withers SB, Sonoyama K, Clancy O, Jeziorska M, Laing I, Yates AP, Pemberton PW, Malik RA, Heagerty AM. Local inflammation and hypoxia abolish the protective anticontractile properties of perivascular fat in obese patients. *Circulation*. 2009;119:1661–1670.
13. Chatterjee TK, Stoll LL, Denning GM, Harrelson A, Blomkalns AL, Idelman G, Rothenberg FG, Neltner B, Romig-Martin SA, Dickson EW, Rudich S, Weintraub NL. Proinflammatory phenotype of perivascular adipocytes: influence of high-fat feeding. *Circ Res*. 2009;104:541–549.
14. Barandier C, Montani JP, Yang Z. Mature adipocytes and perivascular adipose tissue stimulate vascular smooth muscle cell proliferation: effects of aging and obesity. *Am J Physiol Heart Circ Physiol*. 2005;289:H1807–H1813.
15. Schlett CL, Massaro JM, Lehman SJ, Bamberg F, O'Donnell CJ, Fox CS, Hoffmann U. Novel measurements of periaortic adipose tissue in comparison to anthropometric measures of obesity, and abdominal adipose tissue. *Int J Obes (Lond)*. 2009;33:226–232.
16. Deleted in proof.
17. Maurovich-Horvat P, Massaro J, Fox CS, Moselewski F, O'Donnell CJ, Hoffmann U. Comparison of anthropometric, area- and volume-based assessment of abdominal subcutaneous and visceral adipose tissue volumes using multi-detector computed tomography. *Int J Obes (Lond)*. 2007;31:500–506.
18. 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.
19. Meijer WT, Hoes AW, Rutgers D, Bots ML, Hofman A, Grobbee DE. Peripheral arterial disease in the elderly: the Rotterdam Study. *Arterioscler Thromb Vasc Biol*. 1998;18:185–192.
20. 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.
21. 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.
22. Zeymer U, Parhofer KG, Pittrow D, Binz C, Schwertfeger M, Limbourg T, Rother J. Risk factor profile, management and prognosis of patients with peripheral arterial disease with or without coronary artery disease: results of the prospective German REACH registry cohort. *Clin Res Cardiol*. 2009;98:249–256.
23. Golledge J, Leicht A, Crowther RG, Clancy P, Spinks WL, Quigley F. Association of obesity and metabolic syndrome with the severity and outcome of intermittent claudication. *J Vasc Surg*. 2007;45:40–46.
24. Karason K, Peltonen M, Lindroos AK, Sjostrom L, Lonn L, Torgerson JS. Effort-related calf pain in the obese and long-term changes after surgical obesity treatment. *Obes Res*. 2005;13:137–145.
25. Stepp DW, Pollock DM, Frisbee JC. Low-flow vascular remodeling in the metabolic syndrome X. *Am J Physiol Heart Circ Physiol*. 2004;286:H964–H970.

CLINICAL PERSPECTIVE

Central obesity is associated with peripheral arterial disease, suggesting that ectopic fat depots may be associated with localized diseases of the aorta and lower-extremity arteries. We hypothesized that persons with greater amounts of periaortic fat are more likely to have clinical peripheral arterial disease and a low ankle-brachial index. We found that periaortic fat is associated with peripheral arterial disease and low ABI, whereas no association with body mass index, waist circumference, or visceral abdominal fat was observed. Periaortic fat is associated with low ABI and intermittent claudication.

ALCOHOLISM CLINICAL & EXPERIMENTAL RESEARCH

The Mixed Evidence for Brief Intervention in Emergency Departments, Trauma Care Centers, and Inpatient Hospital Settings: What Should We Do?

Craig A. Field, Janette Baird, Richard Saitz, Raul Caetano, and Peter M. Monti

Background: This qualitative review is based on a symposia presented at the 2009 annual conference of the Research Society on Alcoholism (Baird et al., 2009; Field et al., 2009; Monti et al., 2009; Saitz et al., 2009a). The purpose is to describe the mixed evidence supporting brief interventions in the emergency department, trauma care, and in-patient medical care settings; examine potential moderators of treatment outcome in light of the mixed evidence; and identify methods to move the research and practice of brief interventions beyond their current state.

Methods: By drawing upon existing reviews and selected individual studies, we provide examples that reflect the current complexity of research in this area and propose steps for advancing the field.

Results: Emergency departments, inpatient hospital settings, and trauma care settings represent three unique contexts within which brief interventions have been tested. While the general efficacy of brief alcohol interventions in these settings has been recognized, the evidence is increasingly mixed. Recent studies investigating potential moderators of treatment outcomes suggest that a more sophisticated approach to evaluating the effectiveness of brief interventions across varying patient populations is needed to further understand its effectiveness.

Conclusions: Current dissemination efforts represent a significant advance in broadening the base of treatment for alcohol problems by providing an evidence-based intervention in health care settings and should not be curtailed. However, additional research is required to enhance treatment outcomes, refine current practice guidelines, and continue to bridge the gap between science and practice. Given the current state of research, a multisetting clinical trial is recommended to account for potential contextual differences while controlling for study design.

Key Words: Brief Intervention, Emergency Departments, Inpatient Hospital Settings, Review, Trauma Centers.

THE PURPOSE OF this qualitative review of the scientific literature on brief intervention is to (i) describe the mixed evidence supporting brief interventions in the emergency department, trauma care, and inpatient medical care settings; (ii) examine potential moderators of treatment outcome in light of the mixed evidence; and (iii) identify methods to move the research and practice of brief interventions beyond their current state. The intent is not to carry out a systematic review or meta-analysis of screening and brief intervention. Rather, the aim is to draw upon existing reviews and selected individual studies to provide examples that reflect the current complexity of research in this area and propose steps for advancing the field. The review is a product of a symposium presented at the 2009 annual conference of the Research Society on Alcoholism (Baird et al., 2009; Field et al., 2009; Monti et al., 2009; Saitz et al., 2009a,b), which focused on the moderating factors of brief interventions provided to adults in the emergency department, trauma care setting, and inpatient hospital setting.

Emergency departments, inpatient hospital settings, and trauma care settings represent 3 unique contexts within which brief interventions have been tested. In emergency departments, physicians provide care to patients with acute medical and surgical problems including injuries. Emergency department patients with more severe conditions are admitted to the hospital for inpatient care. Inpatient acute care hospital settings are staffed by doctors in medical and surgical specialties and serve a range of patients with diverse acute illnesses and

From the Health Behavior Research and Training Institute (CAF), University of Texas, Austin, Texas; Screening and Brief Intervention (CAF), Trauma Services, University Medical Center Brackenridge, Austin, Texas; Injury Prevention Center (JB), Rhode Island Hospital, Alpert School of Medicine Brown University, Providence, Rhode Island; Research and Education (CARE) Unit (RS), Section of General Internal Medicine, Boston Medical Center, Boston University Schools of Medicine and Public Health, Boston, Massachusetts; UT Southwestern School of Health Professions (RC), Regional Dean and Professor, UT School of Public Health, Dallas, Texas; Center for Alcohol and Addiction Studies (PMM), Brown University, Providence, Rhode Island.

Received for publication January 15, 2010; accepted June 8, 2010.

Reprint requests: Craig Andrew Field, PhD, MPH, University of Texas at Austin, 1717 W 6th Street, Ste 295, Austin, TX 78732; Tel.: +1 512 232 0242; Fax: +1 512 232 0730; E-mail: craig.field@austin.utexas.edu

DOI: 10.1111/j.1530-0277.2010.01297.x

chronic comorbidities. Trauma centers are staffed by trauma surgeons and coexist with emergency departments and other inpatient hospital settings but have a separate accreditation process from emergency departments and hospitals. Trauma centers provide care to severely injured patients most of whom are hospitalized following acute medical treatment for the trauma. This qualitative review distinguishes between these settings given the different contexts in which brief intervention is provided and the differences in patient characteristics presenting to these medical settings. The overarching characteristic these contexts have in common is that they are appropriate settings for opportunistic interventions. Opportunistic interventions are provided to patients with at-risk drinking, alcohol abuse, or dependence who are not seeking treatment for alcohol problems, per se. Brief opportunistic interventions are short, face-to-face conversations regarding drinking, motivation to change, and options for change which are provided during a window of opportunity or potentially teachable moment occasioned by a medical event.

THE CASE FOR BRIEF INTERVENTION

Previous studies have demonstrated the general effectiveness of BI for alcohol problems among patients with trauma. For example, Gentilello and colleagues (1999) showed that adult injured patients with moderate alcohol problems who received BI after being admitted to a Level 1 trauma center decreased alcohol consumption significantly at 12 months, although there was substantial loss to follow-up. Gentilello and colleagues (1999) also observed a statistically nonsignificant reduction in injury requiring treatment in the emergency department or readmission to the trauma center. In an adjusted analysis, Schermer and colleagues (2006) determined that BI significantly reduced arrests for driving under the influence of alcohol (DUI) at 3-year follow-up such that for every 9 interventions provided, there was a reduction in 1 DUI arrest. Finally, Gentilello and colleagues (2005) conducted a cost–benefit analysis based, in part, on the efficacy data of the original clinical trial (Gentilello et al., 1999) and found that for every dollar spent, there was $3.81 saved in direct injury-related medical costs. Therefore, if BI were offered to every eligible injured person in the United States, the resulting savings from health care costs alone would be approximately $1.82 billion annually. On the basis of this evidence, the American College of Surgeons (ACS) mandated that all Level I and Level II trauma centers have a mechanism to identify patients with alcohol problems, and all Level I trauma centers have a mechanism to provide an intervention for patients identified as problem drinkers (Committee on Trauma, 2006). Coinciding with the ACS mandate, the evidence for the efficacy of BI in the trauma care setting became increasingly mixed.

In one of the earliest randomized controlled trials of BI conducted with adult emergency department patients, Longabaugh and colleagues (2001) showed that patients who received the BI in the ED along with a follow-up booster BI showed a significant decrease in alcohol-related negative consequences compared to a standard care group. However, there was no significant difference in reduction between the group receiving a single BI and the standard care group nor were there significant reductions in drinking associated with the BI with booster. More recently, in a quasi-experimental, nonrandomized study, the Academic ED SBIRT (Screening, Brief Intervention and Referral to Treatment in the Emergency Department) Collaborative reported a short-term reduction of 3.25 drinks per week versus controls across 14 ED sites, with 28% of the intervention group no longer exceeding drinking guidelines versus 18% of controls (Academic ED SBIRT Research Collaborative, 2007). In contrast, Daeppen and colleagues (2007), in a randomized trial that included a BI group and 2 control conditions (assessment only and no assessment), found no differential reduction between these 3 groups in terms of alcohol use or health care utilization. Another large randomized trial (D'Onofrio et al., 2008) also showed no significant differences in average volume per week or binge drinking episodes. It is worth noting that both of these negative studies had high follow-up rates. In conclusion, results of research on brief intervention in the emergency department setting are no less ambiguous than those conducted in the trauma center.

A recent review of 14 studies of brief intervention concluded that, in general, there was an effect of brief intervention on reduced alcohol consumption, hazardous use of alcohol, and alcohol-related injuries in comparison with usual emergency department care (Nilsen et al., 2008). However, 5 of the 14 studies found no effect of brief intervention on these outcomes, and even among the studies that found a treatment effect, the BI did not influence similar outcomes. That is, studies that reported a significant effect of BI tended to report either a reduction in alcohol consumption or a reduction in alcohol-related negative consequences, but not both. In almost all of these studies, the control or standard care group also showed decreases in either alcohol consumption or alcohol-related negative consequences that was either coincident with or related to the ED visit.

The general lack of consistency among studies regarding the effectiveness of BI in emergency and trauma departments or the nonspecific effect that BI has on varied alcohol-related outcomes raises questions about the optimal application of BI in these settings. While a substantial amount of research pertaining to the effectiveness of brief intervention has accumulated, inconsistencies in results and the potential impact of various moderators remain unreconciled. Moderators that may be related to patient characteristics, treatment dosing or fidelity factors, or site specific factors (e.g., the context for the intervention) may eliminate, attenuate, or amplify the effects of BI. Future research exploring these aspects may help determine what types of interventions are most effective and what types of patients benefit most. This research would enhance the overall effectiveness of these strategies and/or targeting of interventions to patients who are most likely to benefit.

MODERATORS AFFECTING TREATMENT OUTCOMES: THE USUAL SUSPECTS

One potential explanation for the mixed findings discussed earlier is that, while brief intervention is sometimes effective, certain patient and/or setting characteristics (i.e., context bound effects) are moderators of treatment outcomes following brief interventions. By moderators, we not only mean variation in levels of treatment (e.g., amount or intensity) or in other aspects of treatment (e.g., differences in intervention components or context) but also in participants' predispositions. These cannot be randomized but have to be statistically accounted for as potential sources of variation in treatment outcomes. In a recent systematic review, Nilsen and colleagues (2008) pointed out that moderator analyses of BI effectiveness for reducing alcohol use or alcohol-related negative consequences among trauma or emergency department patients have been empirically evaluated in too few studies to allow for definitive conclusions regarding their influence. However, a number of recent studies have begun to explore potential moderators of treatment outcome following brief intervention in the emergency department.

Monti and colleagues (2009) recently reported differences in treatment outcomes among patients with higher initial motivation to change and those recruited from a trauma unit (in contradistinction to those recruited from emergency department). These patients showed greater reductions in drinking and alcohol-related consequences, but they were not moderators of brief intervention. In their evaluation of moderating effects, Barnett and colleagues (2010) found that drinking at the time of the injury and attribution of the injury to alcohol moderated intervention effects. Participants who were not drinking prior to their injury and those with low or medium attributions and received brief intervention showed lower alcohol use at 12 -month follow-up. In contrast, those who were drinking at the time of their injury and those high in attribution did not show intervention group differences in alcohol use. In contrast, Walton and colleagues (2008) found that patients who received intervention and attributed their injury to alcohol use drank less and had fewer heavy drinking episodes than those who made no attribution of their injury to their alcohol use.

While these studies represent significant strides, they are perhaps no more definitive than studies regarding the main effect of brief intervention. As a result, research on moderators remains essential because, given the mixed findings regarding the main effects of brief intervention, it is essential to delineate subgroups of patients who may differentially benefit from brief intervention. For this review, we have chosen to focus on 2 commonly investigated moderators including severity of alcohol problems and readiness to change. The results of studies examining moderators of behavior change and the mixed findings from these studies further illustrate the challenge of reaching definitive conclusions from existing studies.

Readiness to change alcohol use and severity of alcohol problems have been the most commonly evaluated moderators in studies of BI. Walton and colleagues (2008) found that readiness to change alcohol use and self-efficacy for changing drinking behavior did not moderate effects of BI on alcohol consumption among adult emergency department patients. In contrast, Barnett and colleagues (2010) found that patients in the emergency department with low or medium readiness to change (vs. high) did evidence greater reductions in alcohol use after receiving BI. But, in the inpatient hospital setting, readiness did not appear to moderate the effects of brief intervention (Saitz et al., 2009a,b). Stein and colleagues (2009), in a secondary analysis of the data from the Longabaugh and colleagues (2001) study, found that readiness mediated the effect of treatment on alcohol-related consequences only for those highly motivated to change prior to the intervention but not for those with low pre-intervention motivation. Thus, the observed treatment effect on drinking-related consequences was due, in part, because it helped enhance or maintain readiness to change among those already highly motivated to change. Given that the BI in the Longabaugh study was based on the principles of motivational interviewing, which is intended to increase patient motivation change, the findings of Stein and colleagues (2009) are counterintuitive because ED patients who were less motivated to change (vs. those highly motivated to change) did not benefit from brief intervention. As both Longabaugh and colleagues (2001) and Stein and colleagues (2009) note, a BI delivered in the ED to nontreatment seeking patients may in itself lack the robustness to instigate change without some pre-existing recognition for the need and desire to change alcohol use behaviors among this patient population.

With regard to alcohol severity, a recent study conducted in the trauma care setting by Field and Caetano (2010) determined that brief intervention was effective among admitted injured patients who met DSM-IV criteria for alcohol dependence. This finding is noteworthy because BI has generally been targeted toward nondependent drinkers with the assumption that it was less effective among patients with alcohol dependence. Based on these clinical recommendations, many studies have included indicators of alcohol dependence (e.g., recent history of substance abuse treatment or prior diagnosis) as exclusion criteria. As a result, the findings from Field and Caetano (2010) are particularly important because they may help shed light on recent null findings (most notably, Daeppen et al., 2007; Soderstrom et al., 2007 and Sommers et al., 2006) from well-designed studies that excluded participants who were most likely to meet criteria for alcohol dependence. Mello and colleagues (2008) reporting on the 3-month results from a randomized control trial in which injured emergency department patients received standard care or 2 telephone BI's, found that only those patients who received BI and who had the most severe baseline alcohol scores (as mea-

sured by an Alcohol Use Disorders Identification Test or AUDIT score of 8 or more) showed significant reductions in alcohol-impaired driving. In contrast to the findings of Field and Caetano (2010) and Mello and colleagues (2008), Saitz and colleagues (2009a,b) recently reported that brief intervention was associated with improved outcomes among nondependent drinkers but not among those with alcohol dependence.

These initial studies of potential moderators suggest that BI may not be universally effective across settings and contexts, and that the medical care setting of a hospital emergency department, trauma center, or medical inpatient setting may differ from one another. For example, in the emergency department, there may be numerous interruptions to provide medical care. This may be the reason why Longabaugh found that to produce treatment effects, the participants needed to have received a booster session that took place outside of the emergency department. In contrast, the inpatient setting may present fewer interruptions, but the severity of the patient's medical condition, particularly among patients with trauma, may be a limiting factor to the provision of brief intervention. In addition to the provider's ability to conduct brief intervention, the severity of the medical condition may also influence the patient's receptivity to brief intervention. For example, brief intervention in the trauma department or emergency department may be more effective than inpatient hospital settings because of the saliency of the medical event. Alternatively, the saliency of the event may obviate the need for BI if the medical event itself is sufficient to prompt self-change. In general, differences with regard to findings pertaining to moderating effects may also be a function of different inclusion criteria, intervention conditions, differences in patient population (e.g., age, race/ethnicity, socioeconomic status etc.), or other methodological differences between the studies.

One common criticism of brief intervention studies conducted in the trauma care setting is the apparent inability to account for the impact of the injury event on subsequent drinking behavior. Mello and colleagues (2005) investigated type of injury as a potential moderator and found that brief intervention was effective among those in a motor vehicle collision as opposed to other unintentional or intentional injuries. However, this does not account for the psychological impact of the injury per se that may vary within patients from a particular setting and across treatment contexts. For example, some medical inpatients may have continued to drink despite the development of a symptomatic condition that leads to hospitalization, while others may be provided care for medical problems unassociated with their drinking. Thus, patient receptivity to brief intervention may also vary across hospital settings. Additional efforts to account for individual differences in response to the medical event are particularly important because patient characteristics and context factors may help explain the differential effectiveness of brief intervention across various medical settings.

WHAT IS BRIEF INTERVENTION IN THE MEDICAL SETTING?

While studies in diverse patient populations increase generalizability, the variation in setting and patient population across studies may be an important piece of the puzzle which is often disregarded in systematic reviews, meta-analyses and commentaries or editorials, which attempt to synthesize the results of studies conducted across medical settings as if they were uniform. Nilsen and colleagues (2008) noted that the studies included in their systematic review were conducted in 3 types of medical settings including outpatient emergency departments, inpatient hospitals or trauma centers, and outpatient clinics. This, in and of itself, creates significant problems in terms of deriving general statements about the effectiveness of BI and condensing the results of these studies to yield definitive conclusions. This is due in part because patient characteristics and contextual factors may influence the effectiveness of brief interventions. A meaningful synthesis of current studies also requires greater uniformity in the use of brief intervention, a common approach for identifying patients who receive brief intervention, and increased similarity in both operationalizing and measuring outcomes and consistency in statistical approaches for analyzing those outcomes (Nilsen et al., 2008). Given fundamental differences in these methodologies, we are far from being able to reach definitive conclusions based on the current body of research.

The research on brief intervention in the medical care setting also conflates several distinct types of interventions. For example, brief advice, brief intervention, and brief interventions with multiple visits are all referred to as brief intervention. In a recent meta-analysis of brief intervention in the emergency department setting, 14 studies were identified for inclusion (Moyer et al., 2002). Ten studies evaluated a single session, 8 incorporated principles of motivational interviewing, and 8 provided a handout that included generic advice, personalized advice or feedback based on blood alcohol concentration at the time of admission. In studies that reported the length of the session, the length of the session varied from 5 to 60 minutes. Six studies reported that for the majority of participants, the intervention took place in the emergency department; 2 reported that the majority of sessions took place on an outpatient basis, and 2 studies reported that missed participants were scheduled to return for a visit, and 2 did not indicate the location of the intervention. Beich and colleagues (2003), Emmen and colleagues (2004), and Nilsen and colleagues (2008) reported similar variability in duration, approach, content as well as target population and provider characteristics. As noted by Nilsen and colleagues (2008), the general lack of methodological detail in many studies of brief intervention makes it difficult to discern whether 1 treatment approach or a range of treatment approaches is being evaluated.

While efforts to standardize brief interventions are essential to ongoing research and dissemination, 1 approach for addressing differences in intervention protocols across studies

is to test the underlying theory of brief intervention. Apodaca and Longabaugh (2009), in a recent meta-analysis, evaluated the results of 19 studies and found that limited attention had been given to testing the underlying theory. The most consistent evidence found during their review was that client change talk and client experience of discrepancy were related to better outcomes, and clinician behavior inconsistent with motivational interviewing was related to worse outcomes. Furthermore, the use of a decisional balance exercise showed the strongest association with better outcomes. This initial review of early evidence pertaining to mechanisms of change may provide a starting point for future research investigating the active ingredients and mechanisms of behavior change underlying brief interventions in various medical settings.

RESEARCH ON BRIEF INTERVENTION: NEITHER HERE NOR THERE

Clinical trials are often categorized as either efficacy or effectiveness research that generally focuses on either internal or external validity, respectively. Advancement in clinical research typically proceeds from establishing efficacy in well-controlled clinical trials to testing the generalizability of those findings in effectiveness trials. More recently, alternative conceptualizations to this dichotomy have emerged. Specifically, the terms hybrid research, translational research, practical or pragmatic trials have emerged as similar concepts reflecting the merging of efficacy and effectiveness trials (Woolf, 2008). Brief interventions in the medical setting reflect the shades of gray in hybrid research. Hybrid trials attempt to control some of the parameters that affect outcome while allowing others to vary (Carroll and Rounsaville, 2003). On the one hand, clinical trials of brief intervention in medical settings, including the emergency department and trauma center, have well-established procedures for ensuring adherence to the study protocol such as standardized screening procedures, clearly defined inclusion and exclusion criteria, treatment protocol manuals and ensuring that those conducting assessments are blind to treatment condition. On the other hand, these clinical trials have involved a wide range of patients recruited from diverse settings to evaluate multiple outcomes of interest and diverse approaches to BI. To refer to BI in these settings as either efficacy or effectiveness studies is a misnomer that fails to capture the complexity of conducting clinical trials in these settings.

With regard to behavioral trials including brief intervention in the medical setting, the spectrum may be better be reflected by the framework described by Westfall and colleagues (2007). These authors describe a continuum of research that includes hybrid research as a stage beyond efficacy and effectiveness research but prior to implementation and dissemination. Between these stages of research development, Westfall and colleagues (2007) describe practice-based research, which is necessary before distilled knowledge can effectively be put into practice. Such research focuses on how beneficial and cost effective the treatment is in practice as well as questions about the optimal application and use of the treatment or intervention. This type of research more accurately reflects the challenge of conducting research on brief interventions in medical settings and reaching definitive conclusions regarding their effectiveness. In short, an efficacy study is an unachievable ideal in the medical setting and additional trials to attempt to establish the efficacy of brief intervention in various medical setting without accounting for differences in the treatment context and unique patient characteristics are unlikely to resolve current discrepancies in findings. In addition to testing BI in specific contexts, what is needed is a more sophisticated approach to determining when and how brief intervention is effective in an effort to enhance positive treatment outcomes in various patient populations identified as at-risk drinkers in different treatment contexts.

THEN WHAT SHOULD WE DO? DO NOT THROW THE BABY OUT WITH THE BATHWATER

While not precluding continuing efforts to implement screening and brief intervention, the emergence of mixed evidence should lead to a more careful consideration of the potential strengths and limitations of brief interventions from a research perspective. Given the hybrid nature of the research, highly controlled efficacy trials are not feasible in the emergency department, trauma care, and inpatient medical settings. To date, most studies of brief intervention have evaluated the general effectiveness of brief intervention and have supported the view that some patients do benefit from this approach to reduce at-risk drinking and its consequences. While such studies are an essential first step, similar investigations in the future are unlikely to clarify current ambiguities in the research. For future research to enhance the current knowledge base addressing differences in effects of BI, a systematic approach to evaluating effectiveness is required to continue to support and enhance efforts to effectively implement this evidence-based intervention strategy and to confidently advocate for its widespread dissemination in various medical settings.

As represented in this brief review, research on screening and brief intervention in the medical setting is undoubtedly in a transition period. The current status of research on brief interventions is by no means unique in the investigation of behavioral interventions (see for example, Crits-Christoph, 1997; Kendall, 1998; and Rounsaville and Carroll, 2002). For instance, in Longabaugh and Morgenstern's reviews of cognitive behavioral therapy (CBT), they noted that numerous studies failed to identify which components of CBT accounted for its effectiveness and that many more studies suggested that CBT's effectiveness was limited to specific treatment contexts or specific patient subgroups (Longabaugh and Morgenstern, 1999; Morgenstern and Longabaugh, 2000). Similarly, the most likely explanation for the mixed evidence reported herein is that brief intervention sometimes reduces alcohol use and its associated consequences among certain types of patients in particular treatment contexts.

The mixed results may be a function of the heterogeneity in the patient population, the characteristics of the intervention itself, the setting in which they are provided or differences in research methodology across studies. While single-site randomized trials currently being conducted will undoubtedly prove informative and further enhance the current knowledge base, they are unlikely to disentangle the issues discussed in this qualitative review. Nevertheless, existing research can help determine a path for building on current knowledge to clarify remaining ambiguities. To date, only peripheral or post hoc consideration has been given to the components of the intervention, which are most effective, the therapeutic processes that lead to improved outcomes, the characteristics of patients who respond most favorably to intervention, and contextual factors that may influence the effectiveness of brief intervention. Studies of these factors may help determine if and when these interventions should be offered and thus have important clinical implications that will guide clinical practice and ensure the efficient use of limited resources by targeting patients who are most likely to benefit from intervention and referral for treatment.

A POTENTIAL SOLUTION: ONE PROTOCOL TO RULE THEM ALL

To continue making progress in our understanding of the effectiveness of brief interventions and ensure ongoing dissemination efforts, a more refined approach to its evaluation is required. This review identified 3 core factors which may be influencing our ability to reach definitive conclusions. These include the context in which the intervention is being carried out, the characteristics of patients within these contexts and research methodology including the intervention protocol, measurement, and statistical analysis of data. Traditionally a multisite is considered beneficial, when multiple single-site RCT's are inconclusive because there is sufficient heterogeneity among their results and incongruity in study procedures (Kraemer, 2000). Multisite trials are particularly useful when initial positive findings from well-designed studies are not consistently replicated (Weinberger et al., 2001). A multisite trial would confer numerous benefits above and beyond the current body of single-site studies but would introduce site as a potential confounder to interpreting the findings. Thus, a large multisite trial may not yet be warranted and may not resolve current discrepancies in findings.

An alternative approach to advancing this line of research is to conduct a single-site, multisetting trial of brief intervention. Such a study might further establish the effectiveness of brief intervention, investigate its effectiveness across settings, begin to account for potential moderators unique to the medical setting and differences in patient population, and explore the underlying mechanisms of change. This novel approach would have the advantage of utilizing a single-study methodology, which would standardize the inclusion/exclusion criteria, intervention protocol, assessment procedures, and statistical analysis of the findings. Based on current research, such a study should carefully assess patient characteristics that may influence the effectiveness of brief intervention. Furthermore, the study should also formally test the theoretical mechanism of change underlying brief intervention based on motivational interviewing. Finally, a process evaluation and organizational assessment may help characterize differences between the emergency department, trauma care, and inpatient medical setting, which may influence the delivery of brief intervention and its effectiveness in these settings. Differences between these contexts are perhaps the most difficult and least often studied aspect of conducting behavioral interventions in medical settings. This nontraditional approach to behavioral research may help resolve some of the current discrepancies highlighted in this review by standardizing study procedures across settings and addressing the most likely factors accounting for variations in study findings.

ACKNOWLEDGMENTS

This work was supported by a grant (R01 013824; PI: Caetano) from the National Institute on Alcohol Abuse and Alcoholism to the University of Texas School of Public Health. The lead author acknowledges the support of the NIH Health Disparities Loan Repayment Program funded by the National Center of Minority Health and Health Disparities.

REFERENCES

Academic ED SBIRT Research Collaborative (2007) The impact of screening, brief intervention, and referral for treatment on emergency department patients' alcohol use. Ann Emerg Med 50:699–710.

Apodaca T, Longabaugh R (2009) Mechanisms of change in motivational interviewing: a review and preliminary evaluation of the evidence. Addiction 104:705–715.

Baird J, Nilsen P, Longabaugh R, Mello M, Nirenberg T, Lee C, Woolard R (2009) A review of brief intervention studies conducted in emergency departments: a focus on moderators of treatment outcomes. Alcohol Clin Exp Res 33:269A.

Barnett NP, Apodaca TR, Magill M, Colby SM, Gwaltney C, Rohsenow DJ, Monti PM (2010) Moderators and mediators of two brief interventions for alcohol in the emergency department. Addiction 105:452–465.

Beich A, Thorsen T, Rollnick S (2003) Screening in brief intervention trials targeting excessive drinkers in general practice: systematic review and meta-analysis. Br Med J 327:536–542.

Carroll KM, Rounsaville BJ (2003) Bridging the gap: a hybrid model to link efficacy and effectiveness reaserch in substance abuse treatment. Psychiatr Serv 54:333–339.

Committee on Trauma (2006) Prevention in Resources for Optimal Care of the Injured Patient, pp. 115–120. American College of Surgeons, Chicago, IL.

Crits-Christoph P (1997) Limitations of the dodo bird verdict and the role of clinical trials in psychotherapy research: comment on Wampold et al. (1997). Psychol Bull 122:216–220.

Daeppen J-B (2008) A meta-analysis of brief alcohol interventions in emergency departments: few answers, many questions. Addiction 103: 377–378.

Daeppen J-B, Gaume J, Bady P, Yersin B, Calmes J-M, Givel J-C, Cmel G (2007) Brief alcohol intervention and alcohol assessment do not influence alcohol use in injured patients treated in the emergency department: a randomized clinical trial. Addiction 102:1224–1233.

D'Onofrio G, Pantalon MV, Degutis LC, Flellin DA, Busch SH, Chawarski MC, Owens PH, O'Connor PG (2008) Brief intervention for hazardous and harmful drinkers in the emergency department. Ann Emerg Med 51: 742–750.

Emmen MJ, Schippers GM, Bleijenberg G, Wollersheim H (2004) Effectiveness of opportunistic brief interventions for problem drinking in a general hospital setting: a systematic review. Br Med J 328:318–320.

Field CA, Caetano R (2010) The effectiveness of brief intervention among injured patients with alcohol dependence: who benefits from brief intervention? Drug and Alcohol Dependence. Available at: http://dx.doi.org/10.1016/j.drugalcdep.2009.11.025. Accessed on May 31, 2010.

Field CA, Caetano R, Harris TR, Frankowski R, Roudsari B (2010) Ethnic differences in drinking outcomes following a brief alcohol intervention in the trauma care setting. Addiction 105:62–73.

Field CA, Caetano R, Harris TR, Roudsari B (2009) Understanding ethnic differences in drinking outcomes following a brief alcohol intervention: the role of ethnic matching between patient and provider. Alcohol Clin Exp Res 33:269A.

Gentilello LM, Ebel BE, Wickizer TM, Salkever DS, Rivara FP (2005) Alcohol interventions for trauma patients treated in the emergency departments and hospitals: a cost benefit analysis. Ann Surg 241:541–550.

Gentilello LM, Rivara FP, Donovan DM, Jurkovich GJ, Daranciang E, Dunn CW, Villaveces A, Copass M, Ries RR (1999) Alcohol interventions in a trauma center as a means of reducing the risk of injury recurrence. Ann Surg 230:473–483.

Kendall PC (1998) Empirically supported psychological therapies. J Consult Clin Psychol 66:3–6.

Kraemer HC (2000) Pitfalls of multisite randomized clinical trials of efficacy and effectiveness. Schizophr Bull 26:533–541.

Longabaugh R, Morgenstern J (1999) Cognitive-behavioral coping-skills therapy for alcohol dependence: current studies and future directions. Alcohol Res Health 23:78–96.

Longabaugh R, Woolard RF, Nirenberg TD, Minugh AP, Becker B, Clifford PR, Carty K, Sparadeo F, Gogineni A (2001) Evaluating the effects of a brief motivational intervention for injured drinkers in the emergency department. J Stud Alcohol 62:806–816.

Mello M, Longabaugh R, Baird J, Nirenberg T, Woolard R (2008) DIAL: a telephone brief intervention for high-risk alcohol use with injured emergency department patients. Ann Emerg Med 51:755–764.

Mello M, Nirenberg T, Longabaugh R, Woolard R, Minugh A, Becker B, Baird J, Stein L (2005) Emergency department brief motivational interventions for motor vehicle crash patients. Ann Emerg Med 45:620–625.

Monti P, Colby S, Barnett N, Rohsenow D, Apodaca T, Mello M (2009) Effects of brief alcohol interventions in trauma and emergency departments. Alcohol Clin Exp Res 33:269A.

Morgenstern J, Longabaugh R (2000) Cognitive-behavioral treatment for alcohol dependence: a review of evidence for its hypothesized mechanisms of action. Addiction 95:1475–1490.

Moyer A, Finney JW, Swearingen CE, Vergun P (2002) Brief interventions for alcohol problems: a meta-analytic review of controlled investigations in treatment-seeking populations. Addiction 97:279–292.

Nilsen P, Baird J, Mello MJ, Nirenberg T, Woolard R, Bendtsen P, Longabaugh R (2008) A systematic review of emergency care brief alcohol interventions for injury patients. J Subst Abuse Treat 35:184–201.

Rounsaville BJ, Carroll KM (2002) Commentary on dodo bird revisited: why aren't we dodos yet? Clin Psychol Sci Pract 9:17–20.

Saitz R, Palfai TP, Cheng DM, Horton NJ, Dukes K, Kraemer KL, Roberts MS, Guerriero RT, Samet JH (2009a) Some medical inpatients with unhealthy alcohol use may benefit from brief intervention. Alcohol Clin Exp Res 33:269A.

Saitz R, Palfai TP, Cheng DM, Horton NJ, Dukes K, Kraemer KL, Roberts MS, Guerriero RT, Samet JH (2009b) Some medical inpatients with unhealthy alcohol use may benefit from brief intervention. J Stud Alcohol Drugs 70:426–435.

Schermer CR, Moyers TB, Miller RL, Bloomfield LA (2006) Trauma center brief interventions for alcohol disorders decrease subsequent driving under the influence arrests. J Trauma 60:29–34.

Soderstrom CA, DiClemente CC, Dischinger PC, Hebel JR, McDuff DR, Auman KM, Kufera JA (2007) A controlled trial of brief intervention versus brief advice for at-risk drinking trauma center patients. J Trauma 62:1102–1112.

Sommers MS, Dyehouse JM, Howe SR, Fleming M, Fargo JD, Schafer JC (2006) Effectiveness of brief intervention after alcohol-related vehicular injury: a randomized controlled trial. J Trauma 61:523–533.

Stein LAR, Minugh PA, Longabaugh R, Wirtz P, Baird J, Nirenberg TD, Woolard RF, Carty K, Lee C, Mello M, Becker B, Gogineni A (2009) Readiness to change as a mediator of the effect of a brief motivational intervention on posttreatment alcohol-related consequences of injured emergency department hazardous drinkers. Psychol Addict Behav 23:185–195.

Walton MA, Goldstein AL, Chermack ST, McCammon RJ, Cunningham RM, Barry KL, Blow FC (2008) Brief alcohol intervention in the emergency department: moderators of effectiveness. J Stud Alcohol Drugs 69:550–560.

Weinberger M, Oddone EZ, Henderson WG, Smith DM, Huey J, Giobbie-Hurder A, Feussner JR (2001) Multisite randomized controlled trials in health services research: scientific challenges and operational issues. Med Care 39:627–634.

Westfall JM, Mold J, Fagnan L (2007) Practice-based research-"blue highways" on the NIH roadmap. JAMA 297:403–406.

Woolf SH (2008) The meaning of translation research and why it matters. JAMA 299:211–213.

JRRD Volume 47, Number 8, 2010
Pages 709–718
Journal of Rehabilitation Research & Development

Mental illness-related disparities in length of stay: Algorithm choice influences results

Susan M. Frayne, MD, MPH;[1–3]* Eric Berg, MS;[1–2] Tyson H. Holmes, PhD;[4] Kaajal Laungani, BA;[1] Dan R. Berlowitz, MD, MPH;[5] Donald R. Miller, ScD;[5] Leonard Pogach, MD, MBA;[6] Valerie W. Jackson, MPH;[1–2] Rudolf Moos, PhD[1,3–4]

[1]Center for Health Care Evaluation, Department of Veterans Affairs (VA) Palo Alto Health Care System, Palo Alto, CA; [2]Division of General Internal Medicine, Stanford University, Palo Alto, CA; [3]Center for Primary Care and Outcomes Research, Stanford University, Palo Alto, CA; [4]Department of Psychiatry and Behavioral Sciences, Stanford University, Palo Alto, CA; [5]Center for Health Quality, Outcomes, and Economic Research, Edith Nourse Rogers Memorial Veterans Hospital, Bedford, MA; and Boston University School of Public Health, Boston, MA; [6]Center for Healthcare Knowledge Management, VA New Jersey Health Care System, East Orange, NJ; and University of Medicine and Dentistry of New Jersey-New Jersey Medical School, Newark, NJ

Abstract—Methodological challenges arise when one uses various Veterans Health Administration (VHA) data sources, each created for distinct purposes, to characterize length of stay (LOS). To illustrate this issue, we examined how algorithm choice affects conclusions about mental health condition (MHC)-related differences in LOS for VHA patients with diabetes nationally (n = 784,321). We assembled a record-level database of all fiscal year (FY) 2003 inpatient care. In 10 steps, we sequentially added instances of inpatient care from various VHA sources. We processed databases in three stages, truncating stays at the beginning and end of FY03 and consolidating overlapping stays. For patients with MHCs versus those without MHCs, mean LOS was 17.7 versus 13.6 days, respectively ($p < 0.001$), for the crudest algorithm and 37.2 versus 21.7 days, respectively ($p < 0.001$), for the most refined algorithm. Researchers can improve the quality of data applied to VHA systems redesign by applying methodological considerations raised by this study to inform LOS algorithm choice.

Key words: algorithms, databases, Department of Veterans Affairs, episode of care, healthcare disparities, health services research, human, length of stay, mental disorders, outcome and process assessment, patient discharge, physician's practice patterns, rehabilitation, reproducibility of results, veterans, veterans hospitals.

INTRODUCTION

Health services researchers often use administrative data for characterizing length of stay (LOS) to address a range of objectives. For example, they may examine how LOS (as a dependent variable) varies as a function of patient characteristics (e.g., age, race, insurance status, presence of comorbidity), processes of care (e.g., speed of emergency department response, types of medications administered or interventions applied, discharge protocols, etc.), or institutional characteristics (e.g., teaching hospital, mental health facility, etc.) [1–7]. Alternatively,

Abbreviations: DEpiC = Diabetes Epidemiology Cohort, DSS = Decision Support System, EXT = extended care, FY = fiscal year, ICD-9 = International Classification of Diseases-9th Revision, LOS = length of stay, MHC = mental health condition, OBS = observation, OPAT = outpatient file, VHA = Veterans Health Administration.

***Address all correspondence to Susan M. Frayne, MD, MPH; Center for Health Care Evaluation, 795 Willow Road (152-MPD), Palo Alto, CA 94025; 650-493-5000, ext 23369; fax: 650-617-2690. Email: sfrayne@stanford.edu**

DOI:10.1682/JRRD.2009.08.0112

they may examine LOS as a potential explanatory variable for predicting other outcomes [8] or they may restrict their cohort to patients meeting specific LOS criteria [9]. Furthermore, accurate identification of intervals of inpatient care is required for studies using an episodes-of-care approach [10].

The concept of LOS is simple: time from admission to discharge. However, a number of methodological considerations arise when Veterans Health Administration (VHA) data are used for calculating LOS. First, goals of the project must be carefully considered, because this will influence the algorithm selected. Is the focus on acute or long-term care, on medical-surgical or mental health stays? Is the objective to examine total LOS across multiple years or LOS during a particular interval of study? Second, the algorithm must account for technical, data-quality issues. These include duplicate records, overlapping or sequential inpatient stays, transfers between different inpatient units, and inpatient stays that are recorded in a subsequent year.

Despite that numerous studies focus on LOS, these subtleties of LOS calculation have received little attention. This oversight could have serious implications: algorithm choice can influence conclusions in health services studies [11–13], although to our knowledge this possibility has not been studied in the specific case of LOS. As VHA leadership increasingly seeks to obtain accurate estimates of healthcare costs and use evidence to guide strategic planning decisions, it is critical that the evidence base supporting those decisions be as accurate as possible.

One example of a clinical scenario wherein LOS algorithm choice could influence conclusions is mental health condition (MHC)-related differences in inpatient care use. Prior studies both within and outside the VHA have documented that, compared with patients without MHCs, patients with MHCs tend to use more inpatient care [6,14–19]. Thus, patients with MHCs represent a particularly high-intensity, high-cost group likely to merit special attention by VHA policy makers. However, some characteristics of the way patients with MHC receive inpatient care may make their VHA records disproportionately susceptible to variation in algorithm choice. For example, patients with MHC might be more likely to experience more complex patterns of inpatient care (e.g., transferring between a medical unit and a psychiatric unit during the course of a single hospitalization episode), or to receive care in extended-care settings, where stays can be long and can span multiple fiscal years (FYs). Such factors could potentially influence LOS calculations differently for patients with MHC versus those without MHC.

We used VHA administrative data to examine how application of incrementally more refined algorithms for calculating LOS during 1 year of care affected conclusions about mean LOS in a national cohort of VHA patients with diabetes. Then, as an illustrative example of the practical implications of such methodological decisions, we examined whether the magnitude of observed mental illness-related disparities in mean LOS varied as a function of LOS algorithm applied.

METHODS

Study Context

This work is part of a larger study examining the effect of MHC on processes of outpatient diabetes care in FY03. Because the focus of that study is on outpatient care, we wished to identify (and ultimately exclude from the larger study) patients who were institutionalized (i.e., on inpatient status) for the majority of FY03. Therefore, our goal was to identify, for each patient in our cohort, all days in FY03 during which the patient was on inpatient status (acute care or extended care). We were not seeking to characterize total LOS for the patients in our cohort (which could have spanned multiple years), but only those inpatient days that occurred during FY03. The process of creating our LOS variable and the effect of algorithm choice on conclusions about MHC-related differences in LOS is the focus of the present study.

Subjects

The cohort was drawn from the FY02 Diabetes Epidemiology Cohort (DEpiC), a census of patients with diabetes in VHA nationally. DEpiC is used extensively for VHA epidemiological and health services research [20]. DEpiC identifies patients with diabetes based on the presence of at least one instance of an antiglycemic prescription or at least two instances of a diabetes International Classification of Diseases-9th Revision (ICD-9) code in inpatient or outpatient records. Among the 911,451 FY02 DEpiC members who were veterans, used VHA outpatient care at least once in FY02, and were alive as of the first day of FY03, we selected the 784,321 whose MHC status could be verified, as described next (in subsidiary analyses, we included the full 911,451 subjects, including those with “MHC Possible” status).

Steps to Assemble Raw Record-Level Database of Inpatient Stays

We started by creating a record-level file containing every instance of inpatient care recorded in any inpatient database available in centralized VHA files. We selected only records that contained at least 1 day of inpatient care in FY03. We also deleted duplicate records. In 10 sequential "steps," we pulled all nonduplicate inpatient records containing any FY03 inpatient care for patients in our cohort from the following FY03 files:

Step 1. Bedsection file, which represents acute care hospital stays.

Step 2. OBS (Observation) file, which represents short (e.g., overnight) acute care stays during which the patient is observed regarding the potential need for admission to an acute care bed.

Step 3. EXT (Extended Care) file, which represents long-term care stays (such as rehabilitation stays or nursing home stays).

Step 4. Census file (for Bedsection, OBS, and EXT), which include records for all patients who still held inpatient status on the last day of the FY, and thus for whom a discharge date was not available when the files for that FY were created.

Step 5. Non-VHA file.

Step 6. Fee basis file.

(These latter two files reflect care received outside of VHA but with funding for the care provided by VHA.)

We then searched FY04 and FY05 files for any records that included some FY03 care:

Step 7. Sources 1 through 5, FY04.

Step 8. Fee basis FY04 file (presented separately from other FY04 files to emphasize that fee basis files are more likely to contain "late entry" records from prior years).

Step 9. Sources 1 through 5, FY05.

Step 10. Fee basis FY05 file.

Stages of Processing Record-Level Database of Inpatient Stays

Next, we processed this raw database in sequential "stages." Stage A represented the raw file at any given step. In stage B, we deleted pre-FY03 and post-FY03 care. Specifically, for records with an admission date earlier than the first day of FY03, we deleted any days preceding FY03 (i.e., we modified the record to begin on the first day of FY03), because we were interested in days of care during FY03, not total LOS for the patient across multiple years. Similarly, for records with a discharge date later than the last day of FY03, we modified the record to end on the last day of FY03.

In stage C, we addressed overlapping stays. Several types of overlap were observed, as illustrated in **Figure 1**. In some cases, the entire stay (admission date through discharge date) was contained within the time interval of another record. This might happen, for example, if a patient in a rehabilitation unit was temporarily transferred to an acute care observation bed for an intercurrent illness like pneumonia. If the patient was not formally discharged from the rehabilitation facility prior to the transfer, then the time interval of the short-term stay (appearing in the OBS file) could be bracketed by the interval of the long-term stay (appearing in the EXT file). In other cases an overlap occurred (e.g., the admission date of one record fell between the admission and discharge dates of a subsequent record, or the discharge date of a record fell between the admission and discharge date of a subsequent record). In other cases, contiguous admissions occurred (i.e., the discharge date of one record was the same as the admission date of a subsequent record). For all these overlap cases (which could involve a pair of records or even three or more records), we created a single contiguous episode of FY03 inpatient care by assigning the admission date to be the first admission date in FY03 among the overlapping records and the discharge date to be the last discharge date in FY03 among the overlapping records. The resulting file at step 10, stage C, was our final record-level file of inpatient stays.

Variables

We calculated LOS for each record as the number of days from its start through end dates. At each step/stage, we calculated a cumulative LOS for each patient by adding the record-level LOS for all records included in that step/stage.

To identify patients with MHC, we used the Agency for Health Research and Quality's Clinical Classifications Software (with minor modifications) to generate a list of ICD-9 codes indicating the presence of MHC [21]. A patient was assigned a "Yes" for MHC status if he/she had at least one instance of an MHC ICD-9 code in any inpatient record or outpatient face-to-face clinic visit at baseline (FY01–02) and at least one confirmatory ICD-9 in the study period (FY03). If he/she had no instance of an MHC ICD-9 in FY01 through 03, then he/she was assigned MHC status "No." Otherwise, MHC status was

considered "Possible." That is, the MHC Possible group represents those patients who had an MHC diagnosis in the baseline period or in the study period, but not both. Cases with MHC Possible status were excluded from main analyses; this allowed us to compare LOS in two more sharply defined groups (MHC Yes vs MHC No).

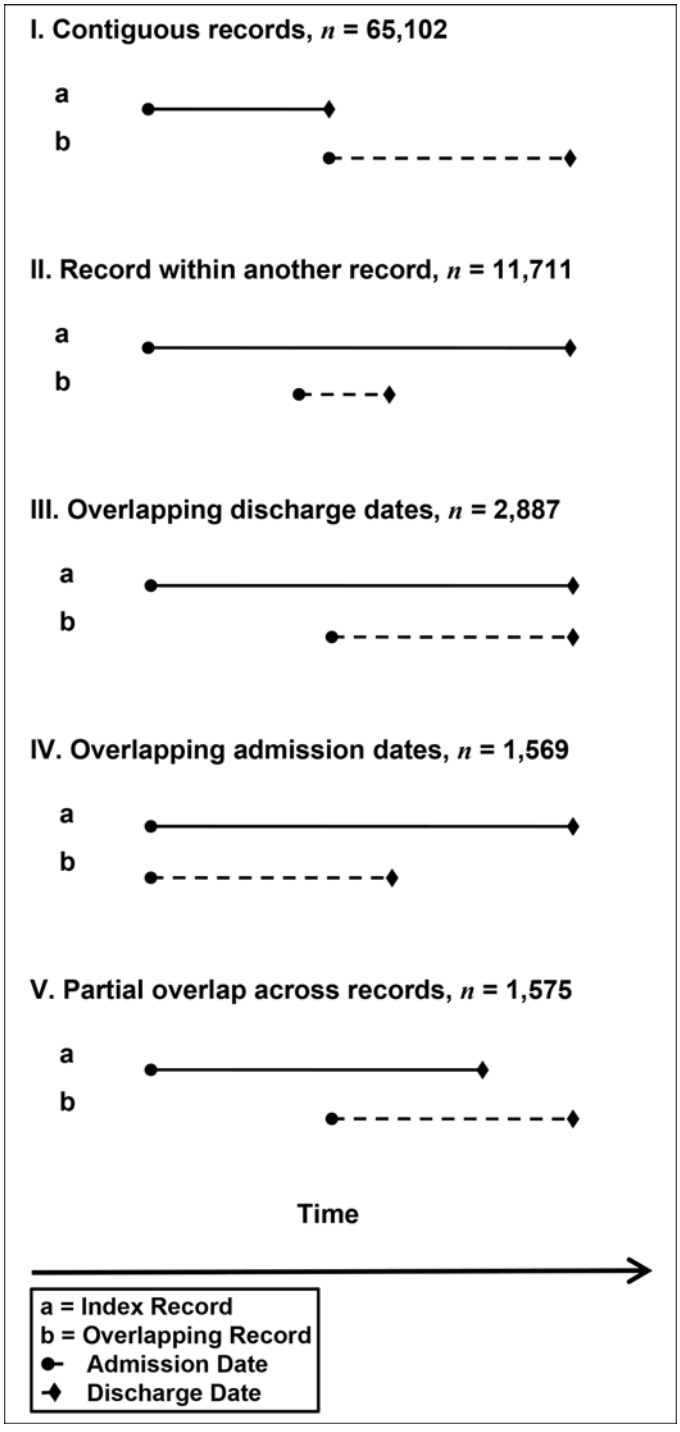


Figure 1.
Patterns I–V of overlap between pairwise records of an individual patient and record-level frequency of each pattern at step 10, stage C.

Analysis

We tabulated the number of records and calculated mean LOS within each cell of a 10 × 3 matrix representing the steps and stages of database development. Next, in each cell, we calculated mean LOS as a function of MHC status. We then calculated the difference (Δ) in mean LOS among patients with MHC versus those without MHC and compared mean LOS for the MHC Yes versus MHC No groups using a two-sample *t*-test. We applied Bonferroni correction for compounding of Type I error across multiple comparisons. Results of hypothesis tests are declared statistically significant for $p < 0.05$ after Bonferroni correction.

RESULTS

Among the 784,321 patients with diabetes in the full cohort, 152,591 were identified as having evidence of an MHC diagnosis (MHC Yes). Among the subset of 92,255 patients who received any inpatient care in FY03 (based on step 10, stage C), 39,452 had MHC Yes. **Table 1** presents the age, sex, Physical Comorbidity Index score (a count from 0–35, developed for case mix adjustment in VHA patients [22–23]), and primary care use in the full cohort and in the subset who used inpatient care, by MHC status.

Table 2 catalogs the number of records and LOS at each step/stage in the database assembly process. The cumulative number of patients who are identified as having received inpatient care in FY03 (based on stage C) increases progressively from step 1 to step 10 (as do the number of records). For example, when the OBS file was added to the Bedsection file, an additional 10,660 records were added for stays that did not perfectly duplicate a Bedsection file stay for that patient. This is expected, because additional evidence of inpatient care is added at each step. More noteworthy is that some steps contribute more records than others.

The number of records does not change at stage B (compared with stage A), because this processing step truncates records (to include only inpatient days during FY03) but does not delete records. However, at stage C (record consolidation), the number of records drops substantially, because overlapping stays are merged into a single, longer stay.

Table 1.
Characteristics of cohort by mental health condition (MHC) status (full cohort and subset who used Veterans Health Administration inpatient care).

Characteristic	Full Cohort, *n* = 784,321		Inpatient Users,* *n* = 92,255	
	MHC Yes	MHC No	MHC Yes	MHC No
n	152,591	631,730	39,452	52,803
Age (years, mean ± SD)	62.1 ± 11.6	69.6 ± 10.3	61.4 ± 11.9	69.1 ± 10.5
Male (%)	96.4	98.5	96.1	98.4
Physical Comorbidity Index (mean ± SD)	3.6 ± 2.4	2.8 ± 2.0	4.6 ± 2.8	4.5 ± 2.7
Used Primary Care in FY03 (%)	93.6	86.8	90.6	91.7

*Inpatient user cohort selected from step 10, stage C.
FY = fiscal year, SD = standard deviation.

Table 2.
Effect of sequential data assembly steps and data cleaning stages on number of patients identified as having received inpatient care and on count of inpatient records and mean length of stay (LOS).

Step	Patients* (*n*)	Records Added in Step (*n*)	Number of Records			LOS (days), Mean ± SD		
			Stage A	Stage B	Stage C	Stage A	Stage B	Stage C
1. Bedsection FY03	77,817	173,707	173,690	173,690	127,566	15.3 ± 21.1	14.8 ± 20.1	14.2 ± 19.4
2. OBS FY03	81,489	10,660	184,350	184,350	135,398	14.9 ± 22.8	14.4 ± 19.8	13.8 ± 19.2
3. EXT FY03	85,198	14,844	199,194	199,194	140,283	27.3 ± 103.8	21.7 ± 37.8	20.7 ± 36.1
4. Census FY03	86,990	6,990	206,184	206,184	143,193	37.6 ± 198.4	26.7 ± 53.1	25.6 ± 51.2
5. Non-VHA FY03	89,135	5,438	211,622	211,622	146,723	37.2 ± 196.2	26.5 ± 52.8	25.4 ± 50.9
6. Fee FY03	90,558	15,107	226,729	226,729	148,721	40.4 ± 198.8	29.8 ± 61.9	27.7 ± 56.2
7. FY04 Records	90,689	5,898	232,627	232,627	148,905	58.3 ± 370.4	35.7 ± 91.2	28.0 ± 57.4
8. Fee FY04	92,068	6,829	239,456	239,456	151,891	58.4 ± 368.5	36.1 ± 93.0	28.3 ± 58.9
9. FY05 Records	92,181	1,569	241,025	241,025	151,980	74.9 ± 518.8	39.4 ± 113.4	28.4 ± 59.3
10. Fee FY05	92,255	293	241,318	241,318	152,157	74.8 ± 518.6	39.4 ± 113.4	28.4 ± 59.3

Note: To create table, we started with step 1 and completed cells across each stage sequentially. Then, for the step 2 analyses, we started with records from steps 1 and 2 and completed cells across each stage sequentially. Analyses for each subsequent step similarly included records from all prior steps. Stages were stage A (original record), stage B (delete days prior to first day of FY03 and after last day of FY03), and stage C (consolidate overlapping stays).
*Reflects cumulative number of patients who received inpatient care in FY03 at each step at stage C. Inpatient records were drawn from patients in analytical cohort (*n* = 784,321).
EXT = extended care, FY = fiscal year, OBS = observation, SD = standard deviation, VHA = Veterans Health Administration.

Consistent with these observations, mean LOS at stage C increased progressively with sequential steps (i.e., as more sources of data were added), except at step 2 (where patients with short OBS stays were added) and at step 5 (where patients with non-VHA stays were added). Similarly, mean LOS decreased progressively with sequential stages. That is, mean LOS decreased from stage A to stage B as non-FY03 days were deleted (which would be relevant to a study like ours that focuses on care received in a single FY). Mean LOS also decreased from stage B to stage C as overlapping days were deleted (which would be relevant to the accuracy of the LOS estimate in any study design). Across the 10 × 3 matrix, mean LOS ranged from 13.8 to 74.9 days.

Table 3 presents LOS by MHC status at every step/stage in the database assembly process. The calculated difference (Δ) in mean LOS between the MHC Yes and the MHC No groups varied markedly by algorithm and was statistically significant ($p < 0.001$) at every step/stage. Correction for multiple comparisons did not statistically affect any findings significantly. As illustrated in **Figure 2**, step 1, Δ = 4.1 at stage A and 3.8 at stage C. In contrast, at step 10, Δ = 57.8 at stage A and 15.5 at stage C ($p < 0.01$ for both between-algorithm comparisons of the values of Δ).

To obtain the LOS in stage C, for each pair of overlapping records, we generated a single record by setting the FY03 admission date as the earliest of the two admission dates and the FY03 discharge date as the latest of the two

Table 3.
Effect of sequential data assembly steps/data cleaning stages on fiscal year (FY) 2003 length of stay (LOS) calculations by mental health condition (MHC) status.

Step	Stage A			Stage B			Stage C		
	MHC Yes	MHC No	Δ	MHC Yes	MHC No	Δ	MHC Yes	MHC No	Δ
1. Bedsection FY03	17.7	13.6	4.1	17.0	13.2	3.8	16.4	12.6	3.8
2. OBS FY03	17.3	13.1	4.2	16.6	12.8	3.8	16.0	12.2	3.8
3. EXT FY03	36.6	20.4	16.2	27.5	17.4	10.1	26.4	16.5	9.9
4. Census FY03	51.6	27.2	24.4	34.0	21.3	12.7	32.7	20.3	12.4
5. Non-VHA FY03	51.0	26.9	24.1	33.7	21.1	12.6	32.4	20.1	12.3
6. Fee FY03	56.1	28.6	27.5	39.1	22.9	16.2	36.2	21.3	14.9
7. FY04 Records	82.1	40.5	41.6	47.1	27.3	19.8	36.7	21.5	15.2
8. Fee FY04	82.3	40.5	41.8	47.8	27.4	20.4	37.1	21.6	15.5
9. FY05 Records	108.0	50.2	57.8	52.5	29.6	22.9	37.3	21.7	15.6
10. Fee FY05	107.9	50.1	57.8	52.4	29.6	22.8	37.2	21.7	15.5

Note: Every difference (Δ) between mean LOS for MHC Yes vs MHC No in this table is statistically significant at $p < 0.001$. Stages were stage A (original record), stage B (delete days prior to first day of FY03 and after last day of FY03), and stage C (consolidate overlapping stays). Two sample *t*-tests were conducted for two key comparisons in this table: comparing within step 1 for stage A vs stage C and within step 10 for stage A vs stage C ($p < 0.01$ for both comparisons).
Δ = mean LOS (MHC Yes) minus mean LOS (MHC No), EXT = extended care, OBS = observation, VHA = Veterans Health Administration.

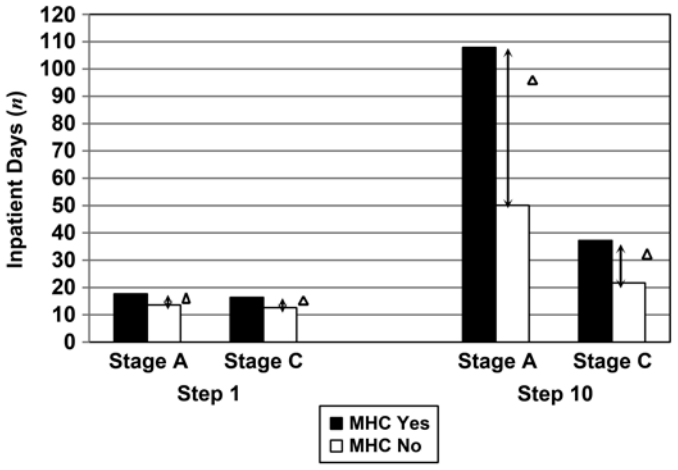


Figure 2.
Effect of sequential data assembly steps and data cleaning stages on fiscal year 2003 number of inpatient days. MHC = mental health condition, Δ = mean length of stay (LOS) (MHC Yes) – mean LOS (MHC No).

discharge dates. We repeated this process iteratively until all pairwise overlaps were addressed. This data processing stage was the most involved, because it needed to account for multiple potential overlap patterns, as illustrated schematically in **Figure 1**. The most common overlap pattern (pattern I) was contiguous records, i.e., where the discharge date of one record was the admission date of the following record. This pattern would happen, for example, if a patient were admitted to one bed section (e.g., to the Psychiatry Department for suicidal ideation) and then transferred to another bed section (e.g., to General Medicine for a hospital-acquired infection). Of note, we used the Bedsection files for these analyses. VA Bedsection files create a new record each time a patient transfers to a different clinical service ("bedsection") during a hospital stay. This is in contrast to the VA Main files, which create a new record for each stay; all contiguous bedsection stays are combined in a single record. Had we used the Main file instead of the Bedsection file, we expect that we would not have encountered this particular form of overlap. Other overlap patterns were also observed, as **Figure 1** shows. Of note, step 10, stage B, yielded LOSs of more than 365 days for 3.2 percent of the MHC Yes group and 1.4 percent of the MHC No group, clearly representing a residual problem with the algorithm; in contrast, no patient had LOS greater than 365 days at stage C. This finding supports the importance of the stage C processing.

In a subsidiary analysis, we found that both the admission and discharge dates fell within FY03 for 91 percent of records at step 10, stage A. In those instances, the full LOS for that episode of care was captured and no truncation was required.

Our main analyses excluded patients who had MHC Possible status (i.e., those patients who had an MHC diagnosis in the baseline period or in the study period, but not both). In another subsidiary analysis (see online **Appendix**), we repeated the main analysis in the initial cohort

(n = 911,451), calculating mean LOS as a function of MHC as a three-way variable (MHC Yes, MHC Possible, MHC No). Mean LOS for the MHC Possible group was consistently intermediate between that for the MHC Yes and MHC No groups. For example, for the MHC Possible group, mean LOS was 16.2 at step 1, stage A; 15.1 at step 1, stage C; 90.3 at step 10, stage A; and 34.4 at step 10, stage C.

DISCUSSION

Choices about what algorithm to use when identifying episodes of inpatient care substantially alter conclusions about the overall intensity of inpatient use and about MHC-related disparities in LOS. Not searching across all appropriate sources of data can lead to failure to capture a substantial amount of inpatient care, thus leading to underestimates of LOS. Decisions about how to process records can likewise influence calculated LOS. While other studies have documented that algorithm choice can influence conclusions drawn from VHA data [11–13], we are not aware of this result having been previously documented for LOS.

Researchers have access to many sources of data about VHA patients' nonambulatory care. Indeed, the large number of sources can bewilder investigators new to VHA administrative data, who may be unsure which files to select. Fortunately, the technical manuals developed by the Department of Veterans Affairs Information Resource Center (available at http://www.virec.research.va.gov/) and the Department of Veterans Affairs Health Economics Resource Center (available at http://www.herc.research.va.gov/) explain these files in detail. Our data provide further empiric information to help guide these decisions. First, our results confirm that adding more data sources identifies more inpatient days. Second, our results indicate that the EXT and Census files are especially important sources of incremental days of inpatient care. Third, our results indicate that adding more data sources also changes conclusions about the magnitude of effect (though not the direction of effect) of MHC on LOS. The step at which this has a particularly pronounced effect is the addition of EXT files, indicating that, compared with patients with no MHC, patients with MHC have disproportionately more frequent or prolonged stays in the long-term care setting.

Investigators using any VHA database need to examine data closely to determine whether data processing steps are necessary. In the case of inpatient files, our data indicate that in addition to the standard procedure of deleting pure duplicate records, investigators must account for overlapping stays (wherein a single day can be counted twice) and, for studies such as ours that focus on a single year of care, to truncate days falling before or after the FY of interest. Such pitfalls could, in some cases, reflect data quality problems, such as a data-entry error in admission or discharge date. However, in many cases, they may not represent deficits in the quality of VHA administrative data, but instead may reflect VHA clinical/administrative record-keeping practices. For example, a single stay could legitimately be recorded in more than one file if these files are used differently. Similarly, a fee basis stay (with the correct admission and discharge dates) could be filed in a subsequent year's records if a delay occurred in receipt of the bill from the outside vendor. Regardless of whether some of these factors represent data quality problems, investigators need to account for them; if not, some patients will have inflated estimates of LOS. Indeed, without such corrections, some patients will appear to be on inpatient status for more than 365 days in a single FY.

While the focus of this study is on the issue of algorithm choice for calculation of LOS, we use MHC-related disparities in LOS as a case study to illustrate what can happen if such issues are not considered. Health services researchers frequently examine disparities in processes and outcomes of care. Historically, interest in disparities related to characteristics like race, sex, and age has been great, but emerging evidence suggests that disparities related to MHC status are also common [9,24]. We demonstrated that the magnitude of MHC-related differences in LOS varied markedly as a function of LOS algorithm. Thus, the methodological issues raised here are not just theoretical: algorithm choice can have marked effects on conclusions in healthcare disparities research.

In the course of conducting analyses for this illustrative example, a subsidiary benefit was that informative findings about associations between MHC status and LOS emerged. Patients with MHC spent more of FY03 on inpatient status than did patients with no MHC; this was a consistent and robust finding across every algorithm examined. This finding is consistent with other studies that have shown heavier use of inpatient services by patients with MHC [6,14–19]. Our study also shows that some types of care (e.g., EXT) are associated with a disproportionately greater MHC effect. Another strength of our approach is that we distinguished between patients

with stronger evidence of MHC (i.e., at least one MHC diagnosis at baseline in FY01–02 and at least one confirmatory MHC diagnosis in the study period, FY03) and patients with less certain (Possible) MHC status (i.e., presence of an MHC diagnosis either at baseline or in the study period, but not both). Our subsidiary analyses provide information about MHC Possible patients, a group that has not been well characterized in prior work. The MHC Possible group is likely heterogeneous and includes patients with an erroneous MHC diagnosis, with transient or resolved MHC, or with less severe MHC, as well as patients who receive part of their care outside the VHA system. Mean LOS for the MHC Possible group consistently fell between the mean LOS observed for the MHC Yes and the MHC No groups.

Interpretation of our findings is subject to several caveats. First, our aim was to calculate total number of days spent on inpatient status during FY03; values should not be interpreted as indicating total LOS across years. However, for 91 percent of records, the patient's complete stay was contained within FY03. Second, we did not use the VHA Decision Support System (DSS) Outpatient (OPAT) file as a data assembly step. In the OPAT file, Stay Type 42, Bedsection 80 refers to nursing home care reimbursed by VHA in any particular month. However, dates of admission and discharge could not be accurately generated from that source. Third, our focus was on VHA use. Depending on an investigator's study question, capturing inpatient days spent in other settings might also be important, such as days identified from Medicare claims data, which can be linked to VHA administrative data [25]. Fourth, because the purpose of our study was to identify periods during which the patient was on nonoutpatient status, our LOS calculations included both acute care and long-term care days. Studies focusing on one or the other setting might need to consider other methodological issues. For example, a patient's stay in a skilled nursing facility could have short gaps (e.g., for a brief acute care stay), which might not be captured with the databases used. Fifth, our main analyses excluded patients whose MHC status could not be ascertained with certainty (MHC Possible), so LOS estimates cannot be generalized to all VHA patients. Subsidiary analyses suggested that these excluded patients had intermediate LOS and that algorithm choice similarly affected LOS calculations for them. Sixth, MHC diagnoses came from ICD-9 diagnosis codes in VHA administrative data rather than from direct assessment of patients' MHC. Given the known problem of underdiagnosis of MHC [26–27], some patients with MHC are likely included in the MHC No group. This would be expected to bias results toward the null.

This study examines methods that should be considered when an algorithm is developed that uses VHA data to calculate LOS. The specific algorithm selected will depend on the research question, such as—

- What types of inpatient care are of interest? For example, is the focus on acute care, extended care, care received on a fee basis outside of VHA or some combination of these sources? If rehabilitative/extended care is the focus, will additional sources (e.g., VHA EXT, fee basis, non-VHA and DSS OPAT files, as well as Medicare or Medicaid files) be queried, and how will multiyear stays be addressed?
- Is the focus on care received in a particular time interval (such as one FY) or on a full episode of inpatient care? If the former, will subsequent years' files be searched for stays recorded in a subsequent FY, and what is the expected incremental benefit versus cost of pulling data from multiple years? If the latter, how many years of data will be searched to identify the complete LOS, which could potentially span many years?
- Is the objective to characterize private sector inpatient care received as well, and if so, should other sources (such as Medicare claims data) be queried?

Careful consideration of these study design issues should yield an algorithm tailored to a particular study's objectives.

CONCLUSIONS

Accounting for the methodological issues raised here should help VHA health services researchers avoid pitfalls in calculation of VHA LOS, such as failure to capture care recorded in more obscure data sources (leading to underestimates of LOS) or duplicate counting of some days of care (leading to overestimates of LOS). This result is expected to support more robust estimates for economic analyses, since inpatient costs contribute disproportionately to total cost of VHA care. This result is also expected to enhance the accuracy of data VHA uses in its evidence-based efforts to redesign its healthcare delivery systems, which aim to improve the quality of care provided to veterans.

ACKNOWLEDGMENTS

Author Contributions:

Study concept and design: S. M. Frayne, E. Berg, T. H. Holmes, R. Moos, D. R. Berlowitz, D. R. Miller, L. Pogach.

Acquisition of data: S. M. Frayne, E. Berg, D. R. Miller.

Analysis and interpretation of data: S. M. Frayne, E. Berg, T. H. Holmes, R. Moos, D. R. Berlowitz, D. R. Miller, L. Pogach, K. Laungani.

Drafting of manuscript: S. M. Frayne, T. H. Holmes, E. Berg.

Critical revision of manuscript for important intellectual content: S. M. Frayne, E. Berg, T. H. Holmes, R. Moos, D. R. Berlowitz, D. R. Miller, L. Pogach, K. Laungani, V. W. Jackson.

Statistical analysis: T. H. Holmes, E. Berg.

Obtained funding: S. M. Frayne.

Administrative, technical, or material support: E. Berg, K. Laungani.

Study supervision: S. M. Frayne.

Financial Disclosures: The authors have indicated that no competing interests exist.

Funding/Support: This material was based on work supported in part by the National Institutes of Health (grant NIDDK 1 R01 DK071202-01) and by VA Health Services Research and Development Service, (grant RCS 90-001).

Additional Contributions: The authors are grateful to Ciaran Phibbs, PhD; Todd Wagner, PhD; and Susan Schmitt, PhD, for their advice and conceptual input during the process of developing the LOS algorithm. The views expressed in this article are those of the authors and do not necessarily represent the views of the VA.

REFERENCES

1. Moos RH, Mertens JR. Patterns of diagnoses, comorbidities, and treatment in late-middle-aged and older affective disorder patients: Comparison of mental health and medical sectors. J Am Geriatr Soc. 1996;44(6):682–88. [PMID: 8642160]
2. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. Med Care. 1998;36(1):8–27. [PMID: 9431328] DOI:10.1097/00005650-199801000-00004
3. Ronis DL, Bates EW, Garfein AJ, Buit BK, Falcon SP, Liberzon I. Longitudinal patterns of care for patients with posttraumatic stress disorder. J Trauma Stress. 1996;9(4): 763–81. [PMID: 8902745] DOI:10.1002/jts.2490090407
4. Saitz R, Ghali WA, Moskowitz MA. The impact of alcohol-related diagnoses on pneumonia outcomes. Arch Intern Med. 1997;157(13):1446–52. [PMID: 9224223] DOI:10.1001/archinte.157.13.1446
5. Cartwright WS, Ingster LM. A patient-based analysis of drug disorder diagnoses in the Medicare population. Health Care Financ Rev. 1993;15(2):89–101. [PMID: 10171899]
6. Ettner SL, Hermann RC. Inpatient psychiatric treatment of elderly Medicare beneficiaries. Psychiatr Serv. 1998;49(9): 1173–79. [PMID: 9735958]
7. Clague JE, Craddock E, Andrew G, Horan MA, Pendleton N. Predictors of outcome following hip fracture. Admission time predicts length of stay and in-hospital mortality. Injury. 2002;33(1):1–6. [PMID: 11879824] DOI:10.1016/S0020-1383(01)00142-5
8. Wigder HN, Johnson C, Shah MR. Length of stay predicts patient and family satisfaction with trauma center services. Am J Emerg Med. 2003;21(7):606–7. [PMID: 14655246] DOI:10.1016/j.ajem.2003.08.019
9. Frayne SM, Halanych JH, Miller DR, Wang F, Lin H, Pogach L, Sharkansky EJ, Keane TM, Skinner KM, Rosen CS, Berlowitz DR. Disparities in diabetes care: Impact of mental illness. Arch Intern Med. 2005;165(22):2631–38. [PMID: 16344421] DOI:10.1001/archinte.165.22.2631
10. Hornbrook MC, Hurtado AV, Johnson RE. Health care episodes: Definition, measurement and use. Med Care Rev. 1985;42(2):163–218. [PMID: 10274864] DOI:10.1177/107755878504200202
11. Frayne SM, Yano EM, Nguyen VQ, Yu W, Ananth L, Chiu VY, Phibbs CS. Gender disparities in Veterans Health Administration care: Importance of accounting for veteran status. Med Care. 2008;46(5):549–53. [PMID: 18438204] DOI:10.1097/MLR.0b013e3181608115
12. Halanych JH, Wang F, Miller DR, Pogach LM, Lin H, Berlowitz DR, Frayne SM. Racial/ethnic differences in diabetes care for older veterans: Accounting for dual health system use changes conclusions. Med Care. 2006;44(5):439–45. [PMID: 16641662] DOI:10.1097/01.mlr.0000207433.70159.23
13. Borzecki AM, Wong AT, Hickey EC, Ash AS, Berlowitz DR. Identifying hypertension-related comorbidities from administrative data: What's the optimal approach? Am J Med Qual. 2004;19(5):201–6. [PMID: 15532912] DOI:10.1177/106286060401900504
14. Ashton CM, Petersen NJ, Wray NP, Yu HJ. The Veterans Affairs medical care system: Hospital and clinic utilization statistics for 1994. Med Care. 1998;36(6):793–803. [PMID: 9630121] DOI:10.1097/00005650-199806000-00003
15. Verbosky LA, Franco KN, Zrull JP. The relationship between depression and length of stay in the general hospital patient. J Clin Psychiatry. 1993;54(5):177–81. [PMID: 8509347]
16. Savoca E. Psychiatric co-morbidity and hospital utilization in the general medical sector. Psychol Med. 1999;29(2):457–64. [PMID: 10218937] DOI:10.1017/S0033291798008071

17. Saravay SM, Steinberg MD, Weinschel B, Pollack S, Alovis N. Psychological comorbidity and length of stay in the general hospital. Am J Psychiatry. 1991;148(3):324–29. [PMID: 1992834]
18. Bressi SK, Marcus SC, Solomon PL. The impact of psychiatric comorbidity on general hospital length of stay. Psychiatr Q. 2006;77(3):203–9. [PMID: 16958003] DOI:10.1007/s11126-006-9007-x
19. Sayers SL, Hanrahan N, Kutney A, Clarke SP, Reis BF, Riegel B. Psychiatric comorbidity and greater hospitalization risk, longer length of stay, and higher hospitalization costs in older adults with heart failure. J Am Geriatr Soc. 2007;55(10):1585–91. [PMID: 17714458] DOI:10.1111/j.1532-5415.2007.01368.x
20. Miller DR, Safford MM, Pogach LM. Who has diabetes? Best estimates of diabetes prevalence in the Department of Veterans Affairs based on computerized patient data. Diabetes Care. 2004;27 Suppl 2:B10–21. [PMID: 15113777] DOI:10.2337/diacare.27.suppl_2.B10
21. Clinical Classifications Software (CCS) for ICD-9-CM [Internet]. Rockville (MD): Healthcare Cost and Utilization Project; 2008. Available from: http://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp.
22. Selim AJ, Fincke G, Ren XS. The comorbidity index. In: Goldfield N, Pine M, Pine J, editors. Measuring and managing health care quality: Procedures, techniques, and protocols. 2nd ed. New York (NY): Aspen; 2002.
23. Selim AJ, Fincke G, Ren XS, Lee A, Rogers WH, Miller DR, Skinner KM, Linzer M, Kazis LE. Comorbidity assessments based on patient report: Results from the Veterans Health Study. J Ambul Care Manage. 2004;27(3):281–95. [PMID: 15287217]
24. Druss BG, Bradford DW, Rosenheck RA, Radford MJ, Krumholz HM. Mental disorders and use of cardiovascular procedures after myocardial infarction. JAMA. 2000;283(4):506–11. [PMID: 10659877] DOI:10.1001/jama.283.4.506
25. Fleming C, Fisher ES, Chang CH, Bubolz TA, Malenka DJ. Studying outcomes and hospital utilization in the elderly. The advantages of a merged data base for Medicare and Veterans Affairs hospitals. Med Care. 1992;30(5):377–91. [PMID: 1583916] DOI:10.1097/00005650-199205000-00001
26. Kimerling R, Ouimette P, Prins A, Nisco P, Lawler C, Cronkite R, Moos RH. Brief report: Utility of a short screening scale for DSM-IV PTSD in primary care. J Gen Intern Med. 2005;21(1):65–67. [PMID: 16423126] DOI:10.1111/j.1525-1497.2005.00292.x
27. Pérez-Stable EJ, Miranda J, Muñoz RF, Ying YW. Depression in medical outpatients. Underrecognition and misdiagnosis. Arch Intern Med. 1990;150(5):1083–88. [PMID: 2184790] DOI:10.1001/archinte.150.5.1083

Submitted for publication August 3, 2009. Accepted in revised form January 13, 2010.

This article and any supplementary material should be cited as follows:
Frayne SM, Berg E, Holmes TH, Laungani K, Berlowitz DR, Miller DR, Pogach L, Jackson VW, Moos R. Mental illness-related disparities in length of stay: Algorithm choice influences results. J Rehabil Res Dev. 2010:47(8): 709–18.
DOI:10.1682/JRRD/2009.08.0112

American Journal of Medical Quality
25(1) 42–50

DOI: 10.1177/1062860609346347
http://ajmq.sagepub.com

Using Administrative Data to Identify Mental Illness: What Approach Is Best?

Susan M. Frayne, MD, MPH,[1] Donald R. Miller, ScD,[2] Erica J. Sharkansky, PhD,[3] Valerie W. Jackson, MPH,[4] Fei Wang, PhD,[2] Jewell H. Halanych, MD, MSc,[5] Dan R. Berlowitz, PhD,[2] Boris Kader, PhD,[6] Craig S. Rosen, PhD,[7] and Terence M. Keane, PhD[8]

Abstract

The authors estimated the validity of algorithms for identification of mental health conditions (MHCs) in administrative data for the 133 068 diabetic patients who used Veterans Health Administration (VHA) nationally in 1998 and responded to the 1999 Large Health Survey of Veteran Enrollees. They compared various algorithms for identification of MHCs from *International Classification of Diseases, 9th Revision* (ICD-9) codes with self-reported depression, posttraumatic stress disorder, or schizophrenia from the survey. Positive predictive value (PPV) and negative predictive value (NPV) for identification of MHC varied by algorithm (0.65-0.86, 0.68-0.77, respectively). PPV was optimized by requiring ≥2 instances of MHC ICD-9 codes or by only accepting codes from mental health visits. NPV was optimized by supplementing VHA data with Medicare data. Findings inform efforts to identify MHC in quality improvement programs that assess health care disparities. When using administrative data in mental health studies, researchers should consider the nature of their research question in choosing algorithms for MHC identification.

Keywords

quality of health care, health services research/methods, algorithms, databases, factual, mental disorders

The burgeoning literature examining quality of care and outcomes for patients with mental health conditions (MHCs) often relies on secondary analyses of administrative databases to answer important questions about this vulnerable population.[1-8] Investigators must take several methodological considerations into account when using these databases to identify patients with MHCs because these databases were developed for clinical and administrative purposes, not for research. Clinical quality improvement programs, likewise, must accurately identify and classify patients with MHCs as part of their efforts to identify subgroups at potential risk for receipt of inferior care; indeed, the Institute of Medicine emphasizes the importance of measuring disparities as a core element of quality assessment.[9] Unfortunately, there is little guidance in the literature about how to accurately identify patients with MHCs from these data sources.

The most widely used typology for classifying MHCs is the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM-IV*), which links explicitly to *International Classification of Diseases, 9th Revision* (ICD-9) codes[10] that are typically available in administrative data sources. When applying these codes, investigators must make various decisions such as which ICD-9 codes to use, the time window over which to search for codes, and

[1]Center for Health Care Evaluation, VA Palo Alto Health Care System, Menlo Park; Division of General Internal Medicine and Center for Primary Care and Outcomes Research, Stanford University, California.
[2]Center for Health Quality, Outcomes & Economic Research, VA Bedford; Boston University School of Public Health, Massachusetts.
[3]National Center for PTSD, VA Boston Healthcare System, Massachusetts.
[4]Center for Health Care Evaluation, VA Palo Alto Health Care System, Menlo Park; Division of General Internal Medicine, Stanford University, California.
[5]University of Alabama, Birmingham.
[6]Center for Health Quality, Outcomes & Economic Research, VA Bedford, Massachusetts.
[7]Center for Health Care Evaluation and National Center for PTSD, VA Palo Alto Health Care System, Menlo Park, California.
[8]National Center for PTSD, VA Boston Healthcare System; Boston University School of Medicine, Massachusetts.

The authors have no financial or other conflicts of interest to disclose.

Corresponding Author:
Susan Frayne, MD, MPH, Center for Health Care Evaluation, 795 Willow Road (152-MPD), Menlo Park, CA 94025
Email: sfrayne@stanford.edu.

the number of times a code must appear for the patient to be considered to have that condition. Empirical data comparing various algorithms for using ICD-9 codes to identify specific MHCs are sparse.[11-13]

Therefore, as part of a national study to examine the effect of mental illness on quality of care for Veterans Health Administration (VHA) patients with diabetes,[14] we developed indicators for any MHC and for specific MHCs (ie, depressive disorders, anxiety, psychosis). To inform our methodological decisions and estimate the validity of our methods, we identified patients from our study who also completed a survey; among them, we compared those with MHCs identified from ICD-9 codes (using various algorithms) with self-report of MHC in the survey. The results of this comparison will be useful to investigators who seek an evidence-based rationale for their choice of methods to identify MHCs using ICD-9 codes.

Methods

Overview

The MEND (MENtal health-Diabetes) study examined disparities in processes and outcomes of diabetes care for diabetes patients with mental illness compared to those without mental illness.[14] Patients were considered to have a MHC if they had at least 1 instance of a MHC diagnosis code in outpatient or inpatient records across a 2-year time window.

For a random subset of patients in the MEND study, we had access to national survey data from VHA Fiscal Year 1999 (FY99), which included questions about a history of diagnosed MHC. Taking advantage of this additional data source, the current study describes our efforts to evaluate how well various ICD-9-based algorithms for identification of MHC predict the presence of self-reported MHCs.

Cohort

The MEND study used data and methods from the VHA Diabetes Epidemiology Cohort (DEpiC), a national database of linked VHA and Medicare data for all VHA diabetes patients since FY98.[15] The MEND cohort included all veterans with diabetes in FY98 who had at least 1 face-to-face outpatient visit at the VHA in FY99 and were alive at the end of FY99 (N = 392059). (DEpiC identifies diabetes patients based on receipt of an antiglycemic prescription or glucose monitoring strips in FY98, or having at least 2 inpatient stays or face-to-face outpatient visits in VHA/Medicare records in FY97-98, associated with diabetes ICD-9 codes 250, 357.2, 362.0, 366.41.) For the analyses presented here, we restricted the MEND cohort to the subset who participated in the 1999 Large Health Survey of Veteran Enrollees (N = 133068).

Data Sources

Data for this study came from 3 existing data sources: VHA's FY99 National Patient Care Database (supplemented with data from FY97-98), Medicare administrative records for the same years, and the VHA Office of Quality & Performance's 1999 Large Health Survey of Veteran Enrollees.[16] The National Patient Care Database contains administrative and clinical outpatient and inpatient records for every VHA patient nationally. Each clinic encounter and each inpatient stay generates a record that includes the clinician-identified ICD-9 diagnosis(es) for care received during that visit. For patients in our cohort, we linked Medicare records from the corresponding years to VHA files using encrypted social security numbers for linkage; sex and date of birth checks confirmed correct linkage.[17]

The 1999 Large Health Survey of Veteran Enrollees, which included questions about a history of diagnosed MHCs, sampled 1.5 million enrollees. Data collection occurred from July 1999 through January 2000. A response rate of 63.1% was achieved[16] using a modified Total Design Method.[18]

Defining MHC, Step 1: Selecting ICD-9 Codes for "Any MHC" and for Specific MHC Categories (The MEND Approach)

The first step was to identify the list of ICD-9 codes corresponding to MHCs. Conceptually, our objective was to identify patients whose clinical presentation would suggest the presence of mental illness to a primary care provider (ie, to clinicians providing diabetes care). We sought to identify a list of MHCs common in the primary care setting and to identify the ICD-9 codes that map to them. To do this, we drew from the conceptual framework developed by a panel of primary care and mental health experts for the American Psychiatric Association's *DSM-IV-Primary Care Edition* (*DSM-IV-PC*).[19,20] This scheme identifies broad clusters of MHCs commonly seen in primary care. It was designed to be used for several purposes, including research, and applies clinical criteria (rather than simply ranges of ICD-9 codes) to select clinically homogeneous clusters of codes. For example, the category "depressive disorder" includes a range of psychiatric conditions such as major depressive disorder, bipolar I disorder currently depressed, adjustment disorder with depressed mood, and depressive disorder not otherwise specified. These distinct

psychiatric conditions all present clinically as falling within the general depressive disorders category. Although the dominant presenting symptoms (eg, depressed mood, anhedonia, irritability) could vary from patient to patient, for a primary care provider, patients in this cluster would all have a somewhat similar clinical appearance. Because primary care providers should be able to recognize classes of psychiatric conditions but might not have the expertise to distinguish between specific psychiatric diagnoses within a class, we concluded that the appropriate level of granularity for grouping MHCs was at the level of condition category rather than at the level of specific diagnosis. The *DSM-IV-PC*'s clinical focus, its orientation toward the primary care setting (where diabetes care is typically delivered), its grouping strategy, and the fact that it explicitly maps to ICD-9 codes made it ideal for our purposes.

To apply this framework to our needs, a panel of 3 practicing internists reviewed the full list of *DSM-IV-PC* codes, eliminating categories (eg, abnormal movements/vocalizations or sexual dysfunction) and individual ICD-9 codes (eg, 305.10—nicotine dependence—or 307.23—Tourette's disorder) that would be seen by primary care providers as medical (rather than psychiatric) in nature. The panel also eliminated codes that reflect a cognitive deficit (eg, dementia, brain injury, learning disabilities), a social problem (eg, V62.89—religious or spiritual problem—or V60.9—housing problem), a condition of childhood (eg, 309.21—separation anxiety disorder—or V62.3—academic problem), or a resolved condition (eg, 291.2—alcohol-induced persisting dementia, implying a prior but not necessarily current substance use disorder). This left us with a set of ICD-9 codes that mapped to 10 specific MHC categories (Table 1): depressive disorder, anxiety, psychotic symptoms, manic symptoms, problematic substance use, dysfunctional personality traits, dissociative symptoms, somatoform symptoms, impulse control disorders, and eating disorders. Patients had *any MHC* if they fell into at least 1 of these condition categories.

Defining MHC, Step 2: Varying the Definition of Any MHC and Specific MHCs

After deciding on the MEND list of ICD-9 codes for MHCs, we developed the requirements for identifying an individual patient as having a MHC. Our base algorithm was algorithm A, which is described below.

Algorithm A. The base algorithm was the presence of at least 1 instance of an ICD-9 code from the MEND list occurring in FY98 or FY99 associated with any VHA inpatient stay or VHA outpatient face-to-face clinical encounter (ie, not telephone encounters, laboratory encounters, or radiology encounters). Many of the common psychiatric conditions we examined tend to be chronic in VHA patients,[21-24] so it is not unreasonable to assume that diagnoses present in one year may be present but not coded in subsequent years. Therefore, we counted codes that were present in the index year (FY99) or the preceding year.

We then sequentially examined 7 alternative algorithms, each of which differed from the base algorithm in 1 dimension.

Algorithm B. Instead of searching FY98-FY99 VHA records for MHC ICD-9 codes, we searched only FY99 records (to address the possibility that some conditions identified in FY98 could have resolved by FY99).

Algorithm C. Instead of requiring 1 or more instances of an MHC ICD-9 code, we required 2 or more instances. This addresses the concern that a single entry of a particular ICD-9 code could reflect a coding error or a "rule out" diagnosis, rather than the true presence of the condition, which is why some authors have required the presence of 2 instances of an ICD-9 code.[12,13]

Algorithm D. Instead of searching all outpatient and inpatient VHA records for MHC, we searched only records from outpatient primary care visits. Because mental illness often presents in a primary care (rather than mental health) setting, many studies focus on primary care diagnoses.

Algorithm E. Instead of searching all outpatient and inpatient VHA records for MHC, we searched only records from outpatient mental health visits. Although not all patients with mental illness receive care in mental health settings, those who do may receive more accurate psychiatric diagnoses than those who visit generalist physicians.

Algorithm F. Instead of using VHA records only, we supplemented VHA records with Medicare outpatient and inpatient administrative records, thereby expanding our scope to include diagnoses made in non-VHA settings.

Algorithm G. Instead of using all ICD-9 codes developed for MEND in step 1, we assembled a Delphi panel of 8 practicing VHA psychologists and psychiatrists to identify a more restricted list of MHC ICD-9 codes. To develop a list of ICD-9 codes expected (based on clinical expertise) to have greater specificity for the presence of MHCs than our full MEND list, the Delphi panel used a 5-point Likert scale to rate the degree of specificity each ICD-9 code possessed for the presence of the symptom category with which it was associated. For example, all Delphi panel members gave a score of 1 (highest specificity) to the ability of ICD-9 code 300.02 (generalized anxiety disorder) to predict the presence of the clinical category "anxiety." In other words, if a clinician coded a visit as having been related to ICD-9 300.02, the Delphi

Table 1. ICD-9 Diagnostic Codes Used to Identify MHCs: The MEND Approach*

Diagnostic Group	ICD-9 Codes
Any MHC	Any of the ICD-9 codes for the 10 categories listed below
Specific categories of MHC	
Depressive disorder	29189, 29284, **29620–29625, 29630–29635, 29650–29655**, 29660–29665, 29689, **3004x, 3090x, 30928, 311xx**
Anxiety	29189, 29289, **30000–30002, 30021–30023, 30029, 3003x, 3083x, 30924, 30981**
Psychosis	2910x, 29281, **2950x-2954x, 2956x, 2957x, 2959x, 29624, 29634, 29644, 29654, 2971x, 2988x, 2989x**
Manic symptoms	29189, 29284, **29600–29605, 29640–29645**, 29660–29665, 2967x, 29680, 29689, 30113
Problematic substance use	**2910x, 2913x, 2915x, 29181**, 29189, **2919x, 2920x**, 29211–29212, **29281**, 29284, 29289, **2929x**, 30300, **30390, 30400, 30410, 30420, 30430, 30440, 30450, 30460, 30490, 30500, 30520, 30530, 30540, 30550, 30560, 30570, 30590**
Dysfunctional personality traits	**3010x, 30120, 30122, 3014x, 30150, 3016x, 3017x, 30181–30183, 3019x**
Dissociative symptoms	**30012–30015, 3006x**
Somatoform symptoms	**30011**, 3007x, **30081**, 30780, 30789
Impulse control disorders	**31230–31234**, 31239
Eating disorders	**3071x, 30750, 30751**

Abbreviations: MHC, Mental health condition; MEND, MENtal health-Diabetes Study; ICD-9, *International Classification of Diseases, 9th Revision*; *DSM-IV-PC*, *Diagnostic and Statistical Manual of Mental Disorders, 4th Edition: Primary Care Version*.
*Following the *DSM-IV-PC* approach, a few ICD-9 codes map to more than 1 category and thus are listed under more than 1 category. This full set of ICD-9 codes was used for algorithms A to F. The subset of ICD-9 codes used in algorithm G (Delphi panel approach) are shown in bold face type.

panel felt that there was a high probability that the clinical presentation of the patient was consistent with anxiety. The Delphi score for each ICD-9 code was the mean of the 8 individual scores. In algorithm G, we examined the effect of using a list of ICD-9 codes that were more restricted than the full MEND list (Delphi score ≤1.1) (Table 1).

Defining MHC, Step 3: Developing a Literature-Based Comparison Algorithm

Algorithm X. To make it easier to put our findings in context with other studies, in our final algorithms, we departed from the MEND approach and developed an algorithm for identifying any MHC (and specific MHCs) using approaches commonly applied in the MHC quality of care literature. Instead of using the ICD-9 codes developed for MEND in step 1, we used a more inclusive set of codes: for any MHC, we used all codes falling within the range 290 to 319 (ie, the complete mental disorders section of the ICD-9 coding guide). This approach has been used by the authors of several major studies in the literature.[25] For the 3 specific MHCs of interest, we used common literature-based approaches to examine an important diagnosis within the depressive disorders category (major depression, defined as ICD-9 296.2x, 296.3x, 311[26,27]), an important diagnosis within the anxiety category (posttraumatic stress disorder [PTSD], defined as ICD-9 309.81), and an important diagnosis within the psychosis category (schizophrenia, defined as ICD-9 295.xx). These narrower diagnoses were selected in part to facilitate comparison with specific self-reported conditions (depression, PTSD, schizophrenia), described next.

Defining Self-Reported MHCs: The Frame of Reference

The 1999 Large Health Survey of Veteran Enrollees included a 15-item checklist that asked about lifetime history of a range of chronic conditions. Participants were asked, "Has a doctor ever told you that you have any of the following?" This stem was followed by a list including 3 MHCs: depression, PTSD, and schizophrenia. A patient responding positively to any of these 3 items was considered to have a self-reported MHC. There is evidence that self-reported medical conditions compare reasonably favorably to medical records, though there is some variability by condition type and less information about psychiatric diagnoses in particular.[28-33]

Analysis Approach

We used ICD-9 codes to calculate the prevalence of "ICD-9-based any MHC" among our full cohort of patients with diabetes who responded to the 1999 Large Health Survey of Veteran Enrollees, sequentially applying our 8 different algorithms (7 MEND-based algorithms and 1 literature-based comparison algorithm). Next, using self-reported MHC as the frame of reference, we calculated positive predictive value (PPV) and negative

Table 2. Demographics, Health Status, and Utilization

	All Diabetics in VHA, N = 392 059	Analytic Cohort, n = 133 068
Sociodemographics		
Age (years), mean (SD)	64.7 (11.3)	66.3 (10.3)
Male (%)	97.9%	98.1%
White (%)	71.3%	76.6%
Health status		
Physical comorbidity, mean (SD)	3.6 (2.6)	3.8 (2.6)
Utilization for FY99		
Outpatient encounters, mean (SD)	17.0 (23.7)	17.4 (22.9)
Outpatient encounters, mental health, mean (SD)	2.3 (11.7)	2.1 (10.8)
Outpatient encounters, primary care, mean (SD)	4.4 (4.6)	4.6 (4.6)
Inpatient stays, mean (SD)	1.9 (1.5)	1.8 (1.5)
Inpatient stays in mental health units, mean (SD)	1.7 (1.4)	1.5 (1.5)

Abbreviations: VHA, Veterans Health Administration; SD, standard deviation; FY99, Fiscal Year 1999.

predictive value (NPV) for each algorithm. PPV is the probability that a patient identified as having a MHC by an ICD-9 code algorithm has independent evidence of having a MHC based on self-report. NPV is the probability that a patient identified as having no MHC by an ICD-9 code algorithm has independent evidence of having no MHC based on self-report.

We then repeated a similar process for 3 specific MHCs (depressive disorders, anxiety disorders, and psychotic disorders), calculating ICD-9-based prevalence of each by applying our 8 algorithms. We then calculated PPV and NPV for ICD-9-based depressive disorders (using self-reported depression as the frame of reference), for ICD-9-based anxiety (using self-reported PTSD as the frame of reference), and for ICD-9-based psychosis (using self-reported schizophrenia as the frame of reference) for each of the 8 algorithms. Note that for algorithm X, we examined ICD-9-based major depression (rather than depressive disorders more generally), ICD-9-based PTSD (rather than anxiety disorders more generally), and ICD-9-based schizophrenia (rather than psychotic disorders more generally).

Results

Our study sample included the 133 068 patients who responded to the 1999 Large Health Survey of Veteran Enrollees from among the total of 392 059 diabetes patients in the MEND cohort in FY99. There were only slight differences between the survey respondents and the overall cohort (Table 2), with survey respondents being older and more often white, having more physical comorbidity and primary care visits and fewer mental health visits. After removing survey respondents who did not answer questions about self-reported MHC, data were available for analysis from 94% of patients (N = 124 716).

Table 3 shows the prevalence, PPV, and NPV for the 8 ICD-9-based algorithms for defining any MHC. Thus, for example, our base algorithm, algorithm A, required that a patient have had at least 1 instance of any MEND ICD-9 code in FY98-99 VHA outpatient or inpatient files to qualify as having a MHC. Algorithm A had a PPV of 0.77 and an NPV of 0.76; this means that, using algorithm A, if we determine that the patient has "ICD-9-based any MHC," then there is a 77% likelihood that he or she has "self-reported MHC" (ie, patient-recalled lifetime history of physician-diagnosed depression, PTSD, and/or schizophrenia). Conversely, if we determine that a patient does not have "ICD-9-based any MHC," then there is a 76% likelihood that he or she has no "self-reported MHC" (ie, no recall of a lifetime history of depression, PTSD, or schizophrenia). We found that PPV could be optimized by requiring at least 2 instances of a MHC ICD-9 code (algorithm C) or by accepting only ICD-9 codes associated with mental health visits (algorithm E). The PPV of these approaches was superior to the PPV of the literature-based approach (algorithm X). In our analyses, NPV was optimized by supplementing VHA data with Medicare data (algorithm F). Estimated prevalence of MHCs also varied by algorithm.

Similarly, Table 4 shows the prevalence, PPV, and NPV for the 8 algorithms for defining the 3 specific MHCs of interest. For all 3 conditions, PPV was optimized by restricting the ICD-9 code list to those identified by the Delphi panel (algorithm G); PPV was also optimized for depression by requiring at least 2 instances of a depression ICD-9 code (algorithm C) and for PTSD by using the literature-based approach (algorithm X). For all 3 conditions, NPV was optimized by supplementing VHA data with Medicare data (algorithm F); the base algorithm also performed well (algorithm A), as did several other

Table 3. Varying Algorithms for Identifying MHCs From Administrative Data: Prevalence, PPV, and NPV Using Self-Reported MHC as the Frame of Reference*

	Any MHC		
	Prevalence	PPV	NPV
Base algorithm, MEND approach:			
Algorithm A: 1+ diagnosis of MHC in FY98 or FY99 from VHA outpatient/inpatient sources, using the list of ICD-9 codes developed for MEND study (from Table 1)	25.1%	0.77	0.76
Modifications of base algorithm, MEND approach:			
Algorithm B: FY99 data only	19.9%	0.82	0.73
Algorithm C: Require 2+ instances of diagnosis	18.8%	0.85	0.73
Algorithm D: Primary care visits only	12.6%	0.77	0.68
Algorithm E: Mental health visits only	17.2%	0.86	0.72
Algorithm F: Supplement VHA data with Medicare data	28.6%	0.74	0.77
Algorithm G: Restricted list of ICD-9 codes from Delphi panel	12.6%	0.80	0.68
Comparison algorithm, literature-based approach			
Algorithm X: literature-based list of ICD-9 codes (full 290–319 range)	34.4%	0.65	0.76

Abbreviations: MHC, mental health condition; PPV, positive predictive value; NPV, negative predictive value; MEND, MENtal-Health-Diabetes Study; FY99, Fiscal Year 1999; VHA, Veterans Health Administration; ICD-9, *International Classification of Diseases, 9th Revision*.
*n = 133 068; for calculation of PPV and NPV, n = 124 716 was used because of missing data for some self-reported MHC items.

algorithms. The generally strong performance of the literature-based approach is not surprising because it compared specific conditions (major depression, PTSD, schizophrenia) with self-reported specific conditions (depression, PTSD, schizophrenia). The frame of reference available to us was thus a better match for the literature-based approach than for the MEND approach, which compared broader categories (depressive disorders, anxiety disorders, psychotic disorders) with the self-reported specific conditions (depression, PTSD, schizophrenia).

Sensitivity and specificity of the various algorithms are available from the author on request.

Discussion

Using a national cohort of veterans with diabetes who use VHA care, we showed that many patients with a diagnosed psychiatric disorder can be detected from VHA administrative records, making this a promising source for quality assessment programs and mental health services research, albeit with several caveats. We demonstrated that choice of algorithm for identification of MHCs influences conclusions about which patients have MHC and, thereby, influences PPV and NPV of MHCs identified from ICD-9 codes.

Although many major studies use administrative data to identify patients with MHC,[1-5] relatively little information is available about how this source can be used most accurately. When using administrative data sources to identify other conditions, such as diabetes[34] and hypertension-related comorbidities,[35] it has been demonstrated that choice of algorithm influences case identification. In the case of psychiatric disorders, though, there is little guidance in the literature. Occasionally, investigators have examined a single algorithm against medical records to identify schizophrenia, depression, or psychiatric disorders as a group.[36-39] However, we are aware of only 1 study that explicitly examined the effect of varying algorithm choice on predictive value for a MHC. Spettell et al[11] compared 2 different algorithms for identification of depression from administrative data (ICD-9 codes and/or antidepressant prescription use) in a managed care organization population, using medical-record-based depression as the frame of reference; PPV varied from 49% to 61%, and NPV varied from 84% to 97%. They did not attempt to vary some of the factors that we varied (types of ICD-9 codes used, source of ICD-9 codes, time interval for ICD-9 code ascertainment, or number of ICD-9 codes used), and they did not examine conditions other than depression. Furthermore, their study was conducted in the private sector; patients with MHCs face fewer barriers (eg, carve-outs, co-pays) in the VHA than in the private sector, and MHC prevalence is particularly high in the VHA.[40-42] Thus, our study is complementary to the prior work by Spettell et al[11] and helps expand the very limited literature on this topic.

Our study has several limitations. First, we compared MHCs in ICD-9 codes against self-reported MHCs because that was the only independent source of information about MHCs available to us. This was by no means an ideal frame of reference because the self-reported data included only 3 psychiatric conditions (depression, PTSD, and schizophrenia), whereas the ICD-9 diagnoses covered depressive disorders, all anxiety disorders (not only PTSD), all psychotic disorders (not only schizophrenia), and multiple other MHCs. Furthermore, the self-reported data were based on lifetime histories of these diagnoses, whereas the ICD-9 code data examined only current conditions. Indeed, given these differences in the way MHC was defined in the 2 sources, it is surprising that we saw as much concordance as we did. Second, the ICD-9 codes

Table 4. Varying Algorithms for Identifying Specific MHCs: Prevalence, PPV, and NPV for Predicting Self-Reported MHC*

	Depressive Disorder			Anxiety			Psychosis		
	Percentage	PPV	NPV	Percentage	PPV	NPV	Percentage	PPV	NPV
Base algorithm, MEND approach:									
Algorithm A: 1+ diagnosis of MHC in FY98 or FY99 from VHA outpatient/inpatient sources, using the list of ICD-9 codes developed for MEND study	16.4%	0.82	0.74	12.6%	0.55	0.90	4.7%	0.46	0.98
Modifications of base algorithm, MEND approach:									
Algorithm B: FY99 data only	12.3%	0.85	0.71	9.4%	0.61	0.89	3.6%	0.53	0.97
Algorithm C: require 2+ instances of diagnosis	11.3%	0.88	0.71	8.4%	0.66	0.89	3.1%	0.60	0.97
Algorithm D: primary care visits only	7.0%	0.82	0.68	4.9%	0.52	0.86	1.3%	0.57	0.96
Algorithm E: mental health visits only	11.4%	0.87	0.71	9.0%	0.65	0.89	3.6%	0.53	0.97
Algorithm F: supplement VHA data with Medicare data	18.5%	0.80	0.75	13.9%	0.52	0.90	6.0%	0.38	0.98
Algorithm G: restricted list of ICD-9 codes from Delphi panel	4.5%	0.90	0.67	3.3%	0.84	0.66	3.5%	0.78	0.66
Comparison algorithm, literature-based approach									
Algorithm X: literature-based list of ICD-9 codes†	14.0%	0.84	0.72	6.0%	0.82	0.88	3.4%	0.58	0.97

Abbreviations: MHC, mental health condition; PPV, positive predictive value; NPV, negative predictive value; MEND, MENtal health-Diabetes Study; FY99, Fiscal Year 1999; VHA, Veterans Health Administration; ICD-9, *International Classification of Diseases, 9th Revision*; PTSD, posttraumatic stress disorder.

*The frame of reference for depressive disorder is self-reported depression; for anxiety, self-reported PTSD; and for psychosis, self-reported schizophrenia. For the calculation of PPV and NPV, participants with missing data for the depression, PTSD, or schizophrenia items on the survey were excluded; the number of participants for these analyses were 122 989, 120 860, and 118 202, respectively.

†Major depression: ICD-9, 296.2x, 296.3x, 311; PTSD, ICD-9 309.81; schizophrenia, ICD-9 295.xx.

identified only diagnosed MHCs; different methods would be needed to identify MHCs not recognized by clinicians[43,44] or MHCs that are recognized but not recorded in the administrative record.[45,46] Third, patients with more severe mental illness could be overrepresented among survey nonrespondents; it is possible that patients who are high functioning enough to reply to the survey are also better able to recall the diagnoses they received previously from a clinician. Fourth, the generalizability of our findings to non-VHA settings (in which the prevalence of MHCs is lower, typically, and mental health carve-outs are common) is unknown. Fifth, the list of ICD-9 codes we developed for the MEND study were designed to capture MHCs that are common in a primary care setting. This list of ICD-9 codes might not translate well to studies focused on care provided in a mental health setting. Finally, it is also noteworthy that although VHA data use ICD-9 codes, a number of institutions have transitioned to the newer ICD-10 coding scheme.

Despite these limitations, our study makes an important contribution to a sparse area in the literature. There have been increasing calls to take advantage of administrative data sources to study important quality of care issues and other health services questions.[47] These sources have significant advantages; for example, they overcome problems of response bias common to studies that seek to actively recruit patients with MHC, reduce human subject burden, and can answer timely questions at relatively low cost. Information about the performance characteristics of these data sets is critical as investigators and quality improvement programs seek to use these sources responsibly.

When applying administrative data to mental health studies, researchers and quality improvement programs should consider whether the nature of their question demands optimization of PPV (eg, by requiring a higher number of ICD-9 code occurrences or by selecting data from mental health clinical sources) or NPV (eg, by searching for diagnoses across more years of data or by supplementing diagnoses recorded in primary care settings with diagnoses recorded in mental health or another data source such as Medicare). Because the limitations of these data sources cannot be fully eliminated with any algorithm, investigators and quality improvement programs should also consider conducting sensitivity analyses in which they vary the algorithm for MHC case identification, so as to bracket their findings by showing how different assumptions affect conclusions.

Acknowledgments

The authors are very grateful to the members of the Delphi Panel (Tina T. Lee, MD, Louis A. Moffett, PhD, Steven E.

Lindley, MD, PhD, Darrah Westrup, PhD, Robert G. Hall, PhD, Gregory A. Leskin, PhD, Matthew Cordova, PhD); to Qing Shao, MD, MS, and William A. West, PhD, who provided technical support for portions of this study; to Rudolf Moos, PhD, who provided valuable comments on an earlier version of this manuscript; and to the Department of Veterans Affairs Office for Quality and Performance for providing the survey data used in this study. The late Mark A. Moskowitz, MD, made key contributions to the early conceptualization of this study.

Authors' Note

Prior presentation: Some of the findings included in this manuscript were part of an oral presentation to the Society of General Internal Medicine Regional Meeting in Stanford, California, in March 2006 an oral presentation to the Society of General Internal Medicine National Meeting in Los Angeles, California, in April 2006; a poster presentation at the Society of General Internal Medicine Regional Meeting in San Francisco, California, in March 2007; and a poster presentation at the American Psychological Association National Meeting in Kansas City, Missouri, in August 2007.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interests with respect to the authorship and/or publication of this article.

Funding

The authors disclosed receipt of the following financial support for the research and/or authorship of this article: This material is based on work supported by the Department of Veterans Affairs Health Services Research & Development, grants IIR 20–041 and RCD 98–312, and by VA Medical Research-Epidemiology. 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 funding sources had no role in the study design, data collection, analysis or interpretation, manuscript preparation, or decision to publish the manuscript.

References

1. Druss BG, Bradford DW, Rosenheck RA, Radford MJ, Krumholz HM. Mental disorders and use of cardiovascular procedures after myocardial infarction. *JAMA.* 2000;283: 506-511.
2. Hermann RC, Ettner SL, Dorwart RA. The influence of psychiatric disorders on patients' ratings of satisfaction with health care. *Med Care.* 1998;36:720-727.
3. Hoff RA, Rosenheck RA, Meterko M, Wilson NJ. Mental illness as a predictor of satisfaction with inpatient care at Veterans Affairs hospitals. *Psychiatr Serv.* 1999;50: 680-685.
4. Druss BG, Rosenheck RA. Use of medical services by veterans with mental disorders. *Psychosomatics.* 1997;38: 451-458.
5. Ashton CM, Petersen NJ, Souchek J, et al. Geographic variations in utilization rates in Veterans Affairs hospitals and clinics. *N Engl J Med.* 1999;340:32-39.
6. Hoge CW, Lesikar SE, Guevara R, et al. Mental disorders among U.S. military personnel in the 1990s: association with high levels of health care utilization and early military attrition. *Am J Psychiatry.* 2002;159: 1576-1583.
7. Seng JS, Clark MK, McCarthy AM, Ronis DL. PTSD and physical comorbidity among women receiving Medicaid: results from service-use data. *J Trauma Stress.* 2006;19:45-56.
8. Smith BM, Weaver FM, Ullrich PM. Prevalence of depression diagnoses and use of antidepressant medications by veterans with spinal cord injury. *Am J Phys Med Rehabil.* 2007;86:662-671.
9. Committee on Quality Health Care in America, Institute of Medicine. *Crossing the Quality Chasm: A New Health System for the 21st Century.* Washington, DC: National Academies Press; 2001.
10. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 4th ed. Washington, DC: American Psychiatric Press; 1994.
11. Spettell CM, Wall TC, Allison J, et al. Identifying physician-recognized depression from administrative data: consequences for quality measurement. *Health Serv Res.* 2003;38:1081-1102.
12. Juster IA, Stensland M, Brauer L, Thuras P. Use of administrative data to identify health plan members with unrecognized bipolar disorder: a retrospective cohort study. *Am J Manag Care.* 2005;11:578-584.
13. Desai MM, Rosenheck RA, Craig TJ. Case-finding for depression among medical outpatients in the Veterans Health Administration. *Med Care.* 2006;44:175-181.
14. Frayne SM, Halanych JH, Miller DR, et al. Disparities in diabetes care: impact of mental illness. *Arch Intern Med.* 2005;165:2631-2638.
15. Miller DR, Safford MM, Pogach LM. Who has diabetes? Best estimates of diabetes prevalence in the Department of Veterans Affairs based on computerized patient data. *Diabetes Care.* 2004;27(suppl 2):B10-B21.
16. Perlin J, Kazis L, Skinner K, et al. *Health Status and Outcomes of Veterans: Physical and Mental Component Summary Scores, Veterans SF36 1999 Large Health Survey of Veteran Enrollees Executive Report.* Washington, DC: Department of Veterans Affairs; 2000.
17. Fleming C, Fisher ES, Chang CH, Bubolz TA, Malenka DJ. Studying outcomes and hospital utilization in the elderly: The advantages of a merged data base for Medicare and Veterans Affairs hospitals. *Med Care.* 1992;30:377-391.
18. Dillman D. *Mail and Telephone Surveys: The Total Design Method.* New York, NY: John Wiley; 1978.
19. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, 4th Edition: Primary Care*

Version (DSM-IV-PC). Washington, DC: American Psychiatric Association; 1995.
20. Pingitore D, Sansone RA. Using DSM-IV primary care version: a guide to psychiatric diagnosis in primary care. *Am Fam Physician.* 1998;58:1347-1352.
21. Peyrot M, Rubin RR. Persistence of depressive symptoms in diabetic adults. *Diabetes Care.* 1999;22:448-452.
22. Klinkman MS, Schwenk TL, Coyne JC. Depression in primary care—more like asthma than appendicitis: the Michigan depression project. *Can J Psychiatry.* 1997;42:966-973.
23. Bremner JD, Southwick SM, Darnell A, Charney DS. Chronic PTSD in Vietnam combat veterans: course of illness and substance abuse. *Am J Psychiatry.* 1996;153:369-375.
24. Daradkeh TK. Stability of psychiatric diagnoses in clinical practice. *Int J Soc Psychiatry.* 1996;42:207-212.
25. Desai MM, Rosenheck RA, Druss BG, Perlin JB. Mental disorders and quality of diabetes care in the Veterans Health Administration. *Am J Psychiatry.* 2002;159:1584-1590.
26. 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.
27. Moos RH, Mertens JR. Patterns of diagnoses, comorbidities, and treatment in late-middle-aged and older affective disorder patients: comparison of mental health and medical sectors. *J Am Geriatr Soc.* 1996;44:682-688.
28. Tisnado DM, Adams JL, Liu H, et al. What is the concordance between the medical record and patient self-report as data sources for ambulatory care? *Med Care.* 2006;44:132-140.
29. Simpson CF, Boyd CM, Carlson MC, Griswold ME, Guralnik JM, Fried LP. Agreement between self-report of disease diagnoses and medical record validation in disabled older women: factors that modify agreement. *J Am Geriatr Soc.* 2004;52:123-127.
30. Skinner KM, Miller DR, Lincoln E, Lee A, Kazis LE. Concordance between respondent self-reports and medical records for chronic conditions: experience from the Veterans Health Study. *J Ambul Care Manage.* 2005;28:102-110.
31. Haapanen N, Miilunpalo S, Pasanen M, Oja P, Vuori I. Agreement between questionnaire data and medical records of chronic diseases in middle-aged and elderly Finnish men and women. *Am J Epidemiol.* 1997;145:762-769.
32. Midthjell K, Holmen J, Bjorndal A, Lund-Larsen G. Is questionnaire information valid in the study of a chronic disease such as diabetes? The Nord-Trondelag diabetes study. *J Epidemiol Community Health.* 1992;46:537-542.
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. 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.
35. Borzecki AM, Wong AT, Hickey EC, Ash AS, Berlowitz DR. Identifying hypertension-related comorbidities from administrative data: what's the optimal approach? *Am J Med Qual.* 2004;19:201-206.
36. Lurie N, Popkin M, Dysken M, Moscovice I, Finch M. Accuracy of diagnoses of schizophrenia in Medicaid claims. *Hosp Community Psychiatry.* 1992;43:69-71.
37. Walkup JT, Boyer CA, Kellermann SL. Reliability of Medicaid claims files for use in psychiatric diagnoses and service delivery. *Adm Policy Ment Health.* 2000;27:129-139.
38. Rawson NS, Malcolm E, D'Arcy C. Reliability of the recording of schizophrenia and depressive disorder in the Saskatchewan health care datafiles. *Soc Psychiatry Psychiatr Epidemiol.* 1997;32:191-199.
39. Kashner TM. Agreement between administrative files and written medical records: a case of the Department of Veterans Affairs. *Med Care.* 1998;36:1324-1336.
40. Hankin CS, Spiro A III, Miller DR, Kazis L. Mental disorders and mental health treatment among U.S. Department of Veterans Affairs outpatients: the Veterans Health Study. *Am J Psychiatry.* 1999;156:1924-1930.
41. Rosenheck R, Greenberg G. *Department of Veterans Affairs National Mental Health Program Performance Monitoring System: Fiscal Year 2001 Report.* West Haven, CT: Northeast Program Evaluation Center, VA Connecticut Healthcare System; 2002.
42. Friedman MJ, Schnurr PP, McDonagh-Coyle A. Posttraumatic stress disorder in the military veteran. *Psychiatr Clin North Am.* 1994;17:265-277.
43. Perez-Stable EJ, Miranda J, Munoz RF, Ying YW. Depression in medical outpatients: underrecognition and misdiagnosis. *Arch Intern Med.* 1990;150:1083-1088.
44. Unutzer J, Schoenbaum M, Druss BG, Katon WJ. Transforming mental health care at the interface with general medicine: report for the presidents commission. *Psychiatr Serv.* 2006;57:37-47.
45. Jencks SF. Recognition of mental distress and diagnosis of mental disorder in primary care. *JAMA.* 1985;253:1903-1907.
46. Rost K, Smith R, Matthews DB, Guise B. The deliberate misdiagnosis of major depression in primary care. *Arch Fam Med.* 1994;3:333-337.
47. Walkup JT, Yanos PT. Psychological research with administrative data sets: an underutilized strategy for mental health services research. *Prof Psychol Res Pr.* 2005;36:551-557.

The Association Between Alcohol Consumption and Prevalent Cardiovascular Diseases Among HIV-Infected and HIV-Uninfected Men

Matthew S. Freiberg, MD, MSc,† Kathleen A. McGinnis, MS,‡ Kevin Kraemer, MD, MSc,* Jeffrey H. Samet, MD, MPH, MA,§ Joseph Conigliaro, MD,‖ R. Curtis Ellison, MD,¶ Kendall Bryant, PhD,# Lewis H. Kuller, MD, DrPH,† and Amy C. Justice, MD, PhD**†† for the VACS Project Team*

Objective: To determine whether alcohol consumption is associated with cardiovascular disease (CVD) among HIV-infected veterans.

Methods: Using established thresholds for alcohol consumption, we analyzed cross-sectional data from 4743 men (51% HIV infected) from the Veterans Aging Cohort Study, a prospective cohort of HIV-infected veterans and demographically similar HIV-uninfected veterans. Using logistic regression, we estimated the odds ratio (OR) for the association between alcohol consumption and prevalent CVD.

Results: Among HIV-infected and HIV-uninfected men, respectively, hazardous drinking (33.2% vs. 30.9%,), alcohol abuse and dependence (20.9% vs. 26.2%), and CVD (14.6% vs. 19.8%) were common. Among HIV-infected men, hazardous drinking [OR = 1.43, 95% confidence interval (CI) = 1.05 to 1.94] and alcohol abuse and dependence (OR = 1.55, 95% CI = 1.07 to 2.23) were associated with a higher prevalence of CVD compared with infrequent and moderate drinking. Among HIV-uninfected men, past drinkers had a higher prevalence of CVD (OR = 1.30, 95% CI = 1.01 to 1.67). For HIV-infected and HIV-uninfected men, traditional risk factors and kidney disease were associated with CVD.

Conclusions: Among HIV-infected men, hazardous drinking and alcohol abuse and dependence were associated with a higher prevalence of CVD compared with infrequent and moderate drinking even after adjusting for traditional CVD risk factors, antiretroviral therapy, and CD4 count.

Key Words: alcohol consumption, alcohol abuse, alcohol dependence, cardiovascular disease, HIV infection, veterans

(*J Acquir Immune Defic Syndr* 2010;53:247–253)

Received for publication February 13, 2009; accepted September 23, 2009.

From the *Division of General Internal Medicine, Department of Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA; †Department of Epidemiology, University of Pittsburgh Graduate School of Public Health, Pittsburgh, PA; ‡Center for Health Equity Research and Promotion, Veterans Affairs Pittsburgh Healthcare System, Pittsburgh, PA; §Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, Boston, MA; ‖Divisions of General Internal Medicine, Lexington Veteran Affairs Medical Center and University of Kentucky Chandler Medical Center, Lexington, KY; ¶Section of Preventive Medicine and Epidemiology, Evans Department of Medicine, Boston University School of Medicine, Boston, MA; #National Institute on Alcohol Abuse and Alcoholism, National Institutes of Health, Rockville, MD; **Veteran Affairs Connecticut Healthcare System, West Haven, CT; and ††Department of Medicine, Yale University School of Medicine, New Haven, CT.

Support for this study came from the Department of Veterans Affairs; the National Institute on Alcohol Abuse and Alcoholism grants; and the National Heart, Lung, and Blood Institute: 2U10AA13566 (Alcohol Associated Outcomes Among HIV+/- Aging Veterans), K23 AA015914 (Alcohol and Coronary Heart Disease in People with HIV), RO1HL095136 (Cardiovascular Disease Mechanisms in HIV Infected and Uninfected Veterans) K24 AA0155674 (Impact of Alcohol Use on HIV Infection-In the United States and Russia), and RO1 AA013304 (Net Lifetime Health Effects of Alcohol Use and Abuse).

Presented at International Society for Biomedical Research on Alcoholism, September 11, 2006, Sydney, Australia; and at the Conference on Retroviruses and Opportunistic Infections, February 25–28, 2007.

The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of the Department of Veteran Affairs.

Correspondence to: Matthew S. Freiberg, MD, MSc, Departments of Medicine and Epidemiology, 230 McKee Place, Suite 600, University of Pittsburgh, Pittsburgh, PA 15213 (e-mail: freibergms@upmc.edu).

INTRODUCTION

With the advent of antiretroviral therapy (ART) and improved survival,[1] alcohol has become an important health issue among HIV-infected adults. It is likely that alcohol is related to several prominent health problems among HIV-infected people including ART adherence,[2] chronic liver disease,[3] possibly HIV disease progression,[4] and cardiovascular disease (CVD).[5] Although the mechanisms for the development of CVD in HIV-infected adults are unknown, ART[6] and perhaps HIV itself[7] are associated with dyslipidemia and increased insulin resistance. In uninfected adults, moderate alcohol consumption is associated with a reduced risk of CVD,[8] improved lipid profiles,[9] increased insulin sensitivity,[10,11] and altered clotting factor profiles.[12] In contrast, hazardous alcohol consumption is associated with hyperlipidemia,[12] incident diabetes,[13] and higher CVD and total mortality rates.[14,15] Although the association between alcohol consumption and CVD risk among uninfected adults is well documented,[16–19] sparse data describe this association among HIV-infected adults. Therefore, the objective of the

present study was to examine the association between alcohol consumption and prevalent CVD among HIV-infected and HIV-uninfected adults from the Veterans Aging Cohort Study (VACS).

METHODS

Veterans

In the present study, we analyzed data on 4743 veterans from the VACS, an observational longitudinal cohort of HIV-infected and HIV-uninfected race, age, and site-matched veterans designed to understand the role of comorbid medical and psychiatric disease in determining clinical outcomes in HIV infection.[20] VACS assesses patients and providers using surveys and electronic medical record review from 8 Veterans Affairs Medical Center Infectious Disease and General Internal Medicine clinics.[20] Data collected included AIDS-defining conditions, comorbidities, health and habits, information about health care provider characteristics, and provider assessments of the participants. A full description of the measures collected and other details regarding the VACS are described elsewhere.[20] From 2002 to 2006, VACS enrolled 6467 participants. Of these, because we know that CVD behaves differently in these groups and due to limited numbers with which to model these differences, women (n = 336) and lifetime abstainers (n = 299) were excluded. Lifetime abstainers were defined as a "No, never" response to, "Have you ever had a drink containing alcohol." Those who had no International Disease Classification 9 (ICD-9) diagnosis code for CVD and were missing self-report CVD information were not included (n = 78). Of the remaining 5762, 669 were excluded for missing alcohol use data and 342 for missing covariate data. The institutional review boards at all locations approved the study, and all veterans provided written informed consent.

Independent Variable

We ascertained infrequent, moderate, and hazardous alcohol consumption using the Alcohol Use Disorders Identification Test (AUDIT).[21] We estimated quantity and frequency of alcohol consumption using the product of the responses to the first 2 questions of the AUDIT: (1) "How often do you have a drink containing alcohol?" and (2) "How many standard drinks do you have on a typical day when you are drinking?" We converted the responses to the first AUDIT question into the following variables: never = 0 times per week; monthly or less = 0.25 times per week; 2–4 times per month = 0.75 times per week; 2–3 times a week = 2.5 times per week; and 4 or more times a week = 4 times per week. For the second AUDIT question, we converted the responses into the following variables: 1 = 1 drink per day; 2 = 2 drinks per day; 3 or 4 = 3.5 drinks per day; 5 or 6 = 5.5 drinks per day; and 7 or more = 7 drinks per day. We calculated weekly drinking as the product of converted responses to questions #1 and #2 (eg, 4 times per week $\times$ 2 drinks per day = 8 drinks per week). Using the question, "When you are drinking, how often do you have 6 or more drinks on one occasion?," we defined a binge drinker as anyone who reported consuming 6 or more drinks on 1 occasion less than monthly or more. Those who responded "never" to consuming 6 or more drinks on 1 occasions were not binge drinkers.

Using this methodology, we categorized alcohol into 3 groups: infrequent and moderate, hazardous, and abuse or dependence. Infrequent and moderate drinkers were combined to form the referent group. Using the National Institute on Alcoholism and Alcohol Abuse guidelines, we defined infrequent or moderate drinking as consuming $\leq$14 drinks per week and no binge drinking. Hazardous drinking was defined as $>$14 drinks per week or binge drinking.[22] Alcohol abuse or dependence was defined using ICD-9 codes based on prior work in the VACS.[23] Importantly, if a participant was a moderate drinker by self-report, but had an ICD-9 code documenting alcohol abuse or dependence, this participant was included in the alcohol abuse and dependent category. We defined past drinkers as those who had consumed $\geq$1 drink in their lifetime but responded "more than 12 months ago" to the question, "When was the last time you had a drink?" As stated earlier, lifetime abstainers were excluded.

Dependent Variable

Our primary outcome variable was prevalent total CVD. We defined CVD using self-reported survey data and VA ICD-9 codes. A participant had CVD if the participant responded yes to 1 of the following 4 separate questions, "Has a doctor ever told you that you had (1) angina or coronary heart disease (CHD), (2) a myocardial infarction, (3) congestive heart failure (CHF), or (4) stroke or transient ischemic attack?" or if the participant had a documented CHD, myocardial infarction, CHF, or stroke event using VA ICD-9 or CPT codes. The complete list of all ICD-9 and CPT codes used in the VACS to define CVD are listed on the VACS website.[24] Using similar methodology, variables were also constructed for CHD, CHF, and stroke, separately.

Covariates

Using VACS patient and provider survey data and Veterans Administration Medical Center pharmacy and laboratory records, we collected data on participant demographics, cardiovascular risk factors, and personal habits. Demographic data included age at VACS study entry and self-reported race/ethnicity (white, black, Hispanic, or other) and education level categorized as either having at least some college education versus high school diploma, general education development (GED), or less education. Cardiovascular risk factors were certain health conditions defined as "yes" response to the question, "Has a doctor ever told you that you had 'high cholesterol, lipids, or triglycerides,' 'diabetes or high blood sugar or sugar,' and 'hypertension or high blood pressure'." Participants were also considered to have "high cholesterol" if there was a documented prescription for an HMG-coreductase inhibitor identified in the pharmacy benefits management database. Current smoking was defined as a yes response to, "Do you now smoke cigarettes?" Body mass index was defined as self-reported weight in kilograms divided by self-reported height in meters squared. HIV-related risk factors included hepatitis C virus (HCV), defined as a positive HCV antibody test, HCV RNA test, or ICD-9 code (070.41, 070.44, 070.51, 070.54 or V02.62), CD4 cell count, and use of and

adherence to ART. We collected data on CD4 cell counts from 180 days before and up to 7 days after the time of enrollment. We defined ART as the use of any antiretroviral medication within the previous 90 days before and up to 7 days after the time of enrollment into the VACS study. For those participants taking ART, nonadherence was defined as having missed at least 1 dose of ART medication in the 4 days before completing the VACS baseline questionnaire. Adherent was not having missed any ART medications in the 4 days before completing the VACS baseline questionnaire. Additional covariates included cocaine use, defined as a yes response to having used cocaine at least once in the past year; self-reported "liver disease or bad liver or cirrhosis;" kidney disease defined as a glomerular filtration rate <30 mL·min^{-1}·1.73 m^{-2}; and regular exercise defined as engaging in regular activities (eg, brisk walking, jogging) long enough to work up a sweat at least 3 times a week.

Statistical Analysis

We obtained descriptive statistics for all variables and assessed the relationship between HIV, alcohol consumption, CVD, and other covariates using *t* tests for continuous variables and χ^2 analysis for categorical variables. We constructed 2 logistic regression models to estimate the odds ratio (OR) for prevalent CVD using level of alcohol use as the main independent variable although adjusting covariates for HIV-infected and HIV-uninfected participants separately. Model 1 adjusted for age, race/ethnicity, and traditional cardiovascular risk factors. Model II adjusted for all covariates in model 1 plus cocaine use, liver disease, kidney disease, exercise, and education. Model 2 for HIV infected also included CD4 cell count and use of and adherence to ART. Secondary analyses also examined the association between level of alcohol use and CVD-specific diagnoses (ie, CHD, CHF, and stroke). Additional analyses were also performed to test separately for the interaction between HIV status and alcohol consumption and the following traditional cardiovascular risk factors: hypertension, hypercholesterolemia, diabetes, and current smoking.

RESULTS

Hazardous alcohol consumption and alcohol abuse or dependence were common among both HIV-infected and HIV-uninfected veterans in the VACS (Table 1). Nearly two-thirds of the veterans were African American. As compared with uninfected veterans, HIV-infected veterans had significantly lower prevalence of several cardiovascular risk factors including hypercholesterolemia, diabetes, hypertension, and mean BMI levels ($P < 0.001$ for all, Table 1). In contrast, HIV-infected veterans had significantly higher prevalence of smoking, HCV, and liver disease ($P < 0.001$ for all). The prevalences of CVD (14.6% vs. 19.8%, $P < 0.001$), CHD (8.6% vs. 14.7%, $P < 0.001$), CHF (4.5% vs. 5.9%, $P = 0.03$), and stroke (5.8 vs. 6.5, $P = 0.30$) were lower among HIV-infected veterans compared with HIV-uninfected veterans.

In both model 1 (adjusted for traditional CVD risk factors) and model 2 (fully adjusted model), hazardous alcohol consumption and alcohol abuse or dependence were associated with an increased prevalence of CVD compared with infrequent or moderate alcohol use for HIV-infected veterans but not HIV-uninfected veterans (Tables 2 and 3). In a model including both HIV-infected and HIV-uninfected veterans (not shown), an interaction term between HIV status and alcohol level was statistically significant ($P = 0.01$). Among HIV-uninfected veterans, past alcohol consumption was associated with a higher prevalence of CVD in both models 1 and 2 (Table 3). For both HIV-infected and HIV-uninfected veterans, traditional risk factors including age, hypercholesterolemia, hypertension, and smoking were associated with a significantly increased prevalence of CVD in models 1 and 2. Kidney disease was also significantly associated with prevalent CVD among HIV-infected and HIV-uninfected veterans (Tables 2 and 3).

When we performed secondary analyses examining separately the interaction between HIV status and traditional cardiovascular risk factors, HIV infection interactions with hypertension ($P = 0.03$), diabetes ($P = 0.04$), and current smoking ($P = 0.01$) were all statistically significant.

We also ran models adjusted for traditional CVD risk factors predicting CHD, CHF, and stroke for HIV-infected and HIV-uninfected veterans. Among HIV-infected veterans, hazardous drinking was statistically significantly associated with CHF (OR = 1.74, 95% CI = 1.04 to 2.91); alcohol abuse or dependence was significantly associated with CHD (OR = 1.67, 95% CI = 1.06 to 2.64) and CHF (OR = 1.99, 95% CI = 1.12 to 3.55); and past drinking was significantly associated with stroke (OR = 1.97, 95% CI = 1.30 to 2.98). Among HIV-uninfected veterans, there were no statistically significant associations between hazardous alcohol consumption and alcohol abuse or dependence and CHD, CHF, or stroke. However, past drinking was statistically significantly associated with stroke (OR = 1.78, 95% CI = 1.24 to 2.54 (data not otherwise shown).

When we examined the association between binge drinking and CVD in a model adjusted for CVD risk factors, there was an increase in the prevalence of CVD among binge drinkers for HIV-infected veterans (OR = 1.30, 95% CI = 1.02 to 1.66). For uninfected veterans, there was no statistically significant increase in CVD among binge drinkers (OR = 1.03, 95% CI = 0.82 to 1.30).

DISCUSSION

In the VACS cohort, among HIV-infected veterans, hazardous drinking and alcohol abuse or dependence were significantly associated with an increased prevalence of CVD as compared with infrequent and moderate drinkers. This association remained significant after adjustment for age, race/ethnicity, traditional CVD risk factors, HCV and liver disease, kidney disease, exercise, education, CD4 count, and adherence to ART. Among HIV-uninfected veterans, past alcohol consumption was associated with a significantly increased prevalence of CVD. In addition to several of the traditional CVD risk factors, renal disease was also significantly associated with a higher prevalence of CVD for both HIV-infected and HIV-uninfected veterans.

Numerous prospective studies among men without HIV report that moderate alcohol consumption is associated with a lower risk of CHD, ischemic stroke, and CVD.[8,14,17,19] In

TABLE 1. Characteristics of the HIV-Infected and HIV-Uninfected Veterans

Characteristics	HIV infected (n = 2422)	Uninfected (n = 2321)	*P*
Demographics			
Mean age yrs (± SD)	49.1 ± 8.7	50.8 ± 9.6	$P < 0.001$
Race/ethnicity (%)			
White	21.8	25.4	$P = 0.003$
Black	65.8	61.4	
Hispanic	8.8	10.2	
Other	3.6	3.0	
>High school education (%)	60.3	58.1	$P = 0.12$
Alcohol consumption (%)			
Past consumption (no alcohol consumption for >1 year)	25.3	27.1	$P = 0.15$
Current infrequent or moderate consumption	45.9	42.9	$P < 0.001$
Current hazardous consumption	33.2	30.9	
Ever alcohol abuse or dependence diagnosis	20.9	26.2	
Cardiovascular risk factors (%)			
Hypercholesterolemia	30.8	41.5	$P < 0.001$
Diabetes	15.2	25.2	$P < 0.001$
Hypertension	32.5	46.8	$P < 0.001$
Current smoking	54.3	47.1	$P < 0.001$
Body mass index	25.2 ± 4.4	28.9 ± 5.6	$P < 0.001$
HIV-related factors			
Hepatitis C positive (%)	46.4	26.4	$P < 0.001$
Mean CD4 count cells/mm^3*	405.0 ± 264.3	—	—
Antiretroviral use†	—	—	—
Not using ART	19.8	—	—
Not Adherent but on ART	26.8	—	—
Adherent on ART	53.5	—	—
Additional covariates (%)			
Cocaine use	23.8	18.1	$P < 0.001$
Liver disease‡	17.2	10.6	$P < 0.001$
Kidney disease (GFR < 30)	2.2	1.1	$P = 0.005$
Regular exercise	54.8	55.6	$P = 0.58$
Type of CVD			
CVD	14.6	19.8	$P < 0.001$
CHD	8.6	14.7	$P < 0.001$
CHF	4.5	5.9	$P = 0.03$
Stroke	5.8	6.5	$P = 0.30$

*Data available only on 2176 veterans.
†Data available only on 2380 veterans.
‡Data available on 4725.

contrast, less data are available focusing on the association between hazardous alcohol consumption, alcohol abuse, and alcohol dependence and CVD. In the present study, among HIV-infected veterans, there was a significant increase in the prevalence of CVD for hazardous drinking and alcohol abuse and dependence as compared with infrequent and moderate drinking. Although the association between alcohol and CVD risk has been thought to be mediated in part by alterations in lipid profiles and levels of clotting factors, prior work in the VACS also demonstrates a temporal and dose–response relationship between alcohol consumption and medication adherence.[25]

Among HIV-uninfected veterans, the association between prevalent CVD and hazardous or abuse and dependence levels of alcohol consumption did not reach statistical significance. In our analyses, interaction terms between HIV status and alcohol consumption, hypertension, diabetes, and current smoking were all significant suggesting that the association between hazardous alcohol consumption, alcohol abuse and dependence, and prevalent CVD is more pronounced among HIV-infected compared with uninfected individuals.

Further, uninfected past drinkers had an increased prevalence of CVD compared to HIV infected past drinkers. It is possible that many of the uninfected veterans who were past drinkers were hazardous alcohol consumers who quit drinking for health-related reasons (ie, sick quitters). Prior research has suggested that "sick quitters" have a higher burden of comorbid disease and thus are at greater risk for CVD.[26] When determining whether HIV-infected individuals in care have a higher or lower risk of CVD compared with uninfected individuals, one must be very clear about the way in which risk

TABLE 2. The Association Between Alcohol Consumption and Other Covariates and CVD Among HIV-Infected Veterans

	Model I CHD Risk Factor Adjusted*	Model II Full Model†
	OR (95% CI), n = 2422	OR (95% CI), n = 2143‡
Demographics		
Age (per 10-year age group)	1.49 (1.29 to 1.73)	1.53 (1.30 to 1.79)
Race		
White	1.0	1.0
Black	0.97 (0.71 to 1.32)	0.95 (0.67 to 1.34)
Hispanic	0.91 (0.54 to 1.53)	0.86 (0.49 to 1.51)
Other	1.86 (0.99 to 3.49)	1.80 (0.92 to 3.52)
>Than high school education	—	1.53 (1.16 to 2.03)
Alcohol consumption		
Infrequent and moderate	1.0	1.0
Hazardous	1.35 (1.01 to 1.79)	1.43 (1.05 to 1.94)
Abuse and dependence	1.51 (1.09 to 2.09)	1.55 (1.07 to 2.23)
Past drinkers (>12 months without a drink) vs. past drinkers (<12 months without a drink or currently drinking)	1.31 (0.99 to 1.71)	1.33 (0.99 to 1.80)
Cardiovascular risk factors		
Hypercholesterolemia	2.37 (1.84 to 3.07)	2.36 (1.77 to 3.13)
Diabetes	1.58 (1.17 to 2.12)	1.71 (1.25 to 2.34)
Hypertension	3.18 (2.45 to 4.12)	2.94 (2.22 to 3.90)
Current smoking	1.80 (1.38 to 2.36)	1.79 (1.33 to 2.41)
Body mass index	0.99 (0.96 to 1.02)	0.99 (0.96 to 1.02)
HIV-related risk factors		
No HCV and no liver disease	—	1.0
No HCV and liver disease	—	1.23 (0.90 to 1.68)
Hepatitis C and no liver disease	—	1.94 (0.99 to 3.80)
Hepatitis C positive and liver disease	—	1.30 (0.88 to 1.92)
Mean CD4 count cells/mm^3‡	—	1.00 (1.00 to 1.00)
Antiretroviral use‡		
Adherent	—	1.00
Therapy and not adherent	—	1.01 (0.74 to 1.38)
No therapy	—	1.05 (0.73 to 1.50)
Other covariates		
Cocaine use	—	1.07 (.76 to 1.52)
Kidney disease (GFR< 30 $mL \cdot min^{-1} \cdot 1.73\ m^{-2}$)	—	2.39 (1.24 to 4.61)
Regular exercise	—	0.81 (0.62 to 1.05)

*CHD risk factor model adjusts for age (in 10-year intervals), race/ethnicity, alcohol consumption, hypercholesterolemia, diabetes, hypertension, current smoking, and body mass index.

†Full Model simultaneously adjusts for age (in 10-year intervals), race, education, alcohol consumption, hypercholesterolemia, diabetes, hypertension, current smoking, body mass index, HCV and liver disease status, cocaine use, kidney disease, exercise, use of and adherence to ART, and CD4 count.

‡Sample size was 2143 for HIV infected because of missing data for CD4 count and ART.

is measured (events being compared) and the comparison population being used. Some prior studies comparing HIV-infected and HIV-uninfected individuals have reported that HIV-infected individuals have a higher prevalence of CVD risk factors[27] or increased Framingham risk score.[28] These studies assume that CVD risk factors are identical between those with and without HIV infection, an assumption that may not be valid. Other studies have reported increased relative risk of incident CVD events or hospitalizations compared with uninfected individuals.[29–32] One of these[30] used a population-based control group. Population-based controls may represent a healthier population as compared with a demographically and behaviorally similar population. VACS used age, race/ethnicity, and clinical site-matched controls. All of the prior studies were conducted in substantially younger populations of both HIV-infected and HIV-uninfected individuals. We observed that the prevalence of CVD was lower among HIV-infected veterans as compared with HIV-uninfected veterans in a population predominated by middle aged and older men. Moreover, most prior studies did not include data on hazardous alcohol consumption, alcohol abuse and dependence, or HCV infection, each of which are important comorbidities among those infected with HIV and can potentially alter cardiovascular risk. Of note, the typical Framingham risk factors (ie, age, hypertension, hypercholesterolemia, diabetes, and smoking) were all significantly associated with CVD in our analyses among HIV-infected veterans. Additionally, kidney disease, as estimated by GFR,

TABLE 3. The Association Between Alcohol Consumption and Other Covariates and CVD Among HIV-Uninfected Veterans

	Model I CHD Risk Factor Adjusted*	Model II Full Model†
	OR (95% CI), n = 2321	OR (95% CI), n = 2321
Demographics		
Age (per year)	1.73 (1.53 to 1.96)	1.74 (1.53 to 1.97)
Race		
White	1.0	1.0
Black	0.69 (0.53 to 0.89)	0.67 (0.51 to 0.87)
Hispanic	1.11 (0.75 to 1.65)	1.09 (0.73 to 1.63)
Other	1.51 (0.82 to 2.78)	1.56 (0.85 to 2.88)
>Than high school education	—	0.93 (0.74 to 1.16)
Alcohol consumption		
Infrequent and moderate	1.0	1.0
Hazardous	0.99 (0.76 to 1.29)	0.97 (0.75 to 1.27)
Abuse and dependence	1.10 (0.81 to 1.49)	0.98 (0.71 to 1.35)
Past drinkers (>12 months without a drink) vs. past drinkers (<12 months without a drink or currently drinking)	1.27 (0.99 to 1.62)	1.30 (1.01 to 1.67)
Cardiovascular risk factors		
Hypercholesterolemia	2.83 (2.23 to 3.60)	2.88 (2.26 to 3.68)
Diabetes	1.07 (0.83 to 1.37)	1.04 (0.80 to 1.34)
Hypertension	2.24 (1.76 to 2.85)	2.26 (1.77 to 2.88)
Current smoking	1.37 (1.07 to 1.76)	1.33 (1.03 to 1.73)
Body mass index	1.00 (0.98 to 1.02)	1.00 (0.98 to 1.02)
HIV-related risk factors		
No HCV and no liver disease	—	1.0
No HCV and liver disease	—	0.90 (0.65 to 1.24)
Hepatitis C and no liver disease	—	1.52 (0.80 to 2.87)
Hepatitis C positive and liver disease	—	1.14 (0.76 to 1.72)
Other covariates		
Cocaine use	—	1.46 (1.04 to 2.05)
Kidney disease (GFR<30 $mL \cdot min^{-1} \cdot 1.73\ m^{-2}$)	—	2.42 (1.03 to 5.72)
Regular exercise	—	1.08 (0.86 to 1.36)

*CHD risk factor model adjusts for age (in 10-year intervals), race, alcohol consumption, hypercholesterolemia, diabetes, hypertension, current smoking, and body mass index.
†Full model simultaneously adjusts for age (in 10-year intervals), race/ethnicity, education, alcohol consumption, hypercholesterolemia, diabetes, hypertension, current smoking, body mass index, HCV and liver disease status, cocaine use, kidney disease, and exercise.

was also significantly associated with CVD. This result is consistent with prior findings among HIV-uninfected people.[33]

The present study has several limitations that warrant comment. As this study is cross sectional, we cannot comment on cause and effect with regard to alcohol consumption and the risk of CVD. Further, associations with prevalent CVD may differ from those with incident disease in HIV because at least some risk factors (hyperlipidemia and glucose intolerance) increase with exposure to antiretroviral treatment. As there were only men in the present study, our findings may not be generalizable to women. As several variables in the analyses involved self-reported data, there is the possibility of nondifferential misclassification. Further, there may have been some nondifferential misclassification among those who were HCV antibody positive but without HCV RNA because HCV infection spontaneously resolves in 10%–15%.[34] In addition, there is the possibility of misclassification among the HIV-uninfected VACS participants. However, the possibility of seroconversion of an HIV-uninfected participant is unlikely. In the prior decade of conducting the VACS studies, less than 0.2% of the patients classified as HIV uninfected have been subsequently identified as infected. It would also be helpful to have had more complete data to differentiate past drinkers into those who quit for health-related reasons versus nonhealth-related reasons. However, we did include alcohol diagnoses which helped to further categorize current drinking, particularly among those who were currently infrequent or moderate drinkers. Finally, we found significant interaction terms suggesting that risk factors for CVD demonstrate different associations with CVD among those infected with HIV compared with uninfected individuals. These findings underscore the importance of studying actual clinical events rather than risk factors if we are to gain a better understanding of CVD risk among those with HIV infection.

In conclusion, hazardous alcohol consumption and alcohol abuse or dependence were associated with an increased prevalence of CVD among HIV-infected veterans compared with infrequent and moderate alcohol consumption. This association persisted even after adjustment for traditional CVD risk factors, HIV-related risk factors including HCV, use

of and adherence to ART, and CD4 count. This association did not reach significance among uninfected demographically similar comparators suggesting that the effect of alcohol may be more pronounced among those infected with HIV.

REFERENCES

1. Palella FJ Jr, Delaney KM, Moorman AC, et al. Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. HIV Outpatient Study Investigators. *N Engl J Med*. 1998;338: 853–860.
2. Samet JH, Horton NJ, Meli S, et al. Alcohol consumption and antiretroviral adherence among HIV infected persons with alcohol problems. *Alcohol Clin Exp Res*. 2004;28:572–577.
3. Fultz SL, Justice AC, Butt AA, et al. Testing, referral, and treatment patterns for hepatitis C virus coinfection in a cohort of veterans with human immunodeficiency virus infection. *Clin Infect Dis*. 2003;36: 1039–1046.
4. Samet JH, Cheng DM, Libman H, et al. Alcohol consumption and HIV disease progression. *J Acquir Immune Defic Syndr*. 2007;46:194–199.
5. Hadigan C, Jeste S, Anderson EJ, et al. Modifiable dietary habits and their relation to metabolic abnormalities in men and women with human immunodeficiency virus infection and fat redistribution. *Clin Infect Dis*. 2001;33:710–717.
6. Friis-Moller N, Sabin CA, Weber R, et al. Combination antiretroviral therapy and the risk of myocardial infarction. *N Engl J Med*. 2003;349: 1993–2003.
7. Paton P, Tabib A, Loire R, et al. Coronary artery lesions and human immunodeficiency virus infection. *Res Virol*. 1993;144:225–231.
8. Corrao G, Rubbiati L, Bagnardi V, et al. Alcohol and coronary heart disease: a meta-analysis. *Addiction*. 2000;95:1505–1523.
9. Gaziano JM, Buring JE, Breslow JL, et al. Moderate alcohol intake, increased levels of high-density lipoprotein and its subfractions, and decreased risk of myocardial infarction. *N Engl J Med*. 1993;329: 1829–1834.
10. Davies MJ, Baer DJ, Judd JT, et al. Effects of moderate alcohol intake on fasting insulin and glucose concentrations and insulin sensitivity in postmenopausal women: a randomized controlled trial. *JAMA*. 2002;287: 2559–2562.
11. Rimm EB, Chan J, Stampfer MJ, et al. Prospective study of cigarette smoking, alcohol use, and the risk of diabetes in men. *BMJ*. 1995;310: 555–559.
12. Rimm EB, Williams P, Fosher K, et al. Moderate alcohol intake and lower risk of coronary heart disease: meta-analysis of effects on lipids and haemostatic factors. *BMJ*. 1999;319:1523–1528.
13. Carlsson S, Hammar N, Efendic S, et al. Alcohol consumption, Type 2 diabetes mellitus and impaired glucose tolerance in middle-aged Swedish men. *Diabet Med*. 2000;17:776–781.
14. McElduff P, Dobson AJ. How much alcohol and how often? Population based case-control study of alcohol consumption and risk of a major coronary event. *BMJ*. 1997;314:1159–1164.
15. Malyutina S, Bobak M, Kurilovitch S, et al. Relation between heavy and binge drinking and all-cause and cardiovascular mortality in Novosibirsk, Russia: a prospective cohort study. *Lancet*. 2002;360:1448–1454.
16. Mukamal KJ, Chung H, Jenny NS, et al. Alcohol use and risk of ischemic stroke among older adults: the cardiovascular health study. *Stroke*. 2005; 36:1830–1834.
17. Mukamal KJ, Conigrave KM, Mittleman MA, et al. Roles of drinking pattern and type of alcohol consumed in coronary heart disease in men. *N Engl J Med*. 2003;348:109–118.
18. Beulens JW, Rimm EB, Ascherio A, et al. Alcohol consumption and risk for coronary heart disease among men with hypertension. *Ann Intern Med*. 2007;146:10–19.
19. Mukamal KJ, Ascherio A, Mittleman MA, et al. Alcohol and risk for ischemic stroke in men: the role of drinking patterns and usual beverage. *Ann Intern Med*. 2005;142:11–19.
20. Justice AC, Dombrowski E, Conigliaro J, et al. Veterans Aging Cohort Study (VACS): overview and description. *Med Care*. 2006;44(8 Suppl 2): S13–S24.
21. Saunders JB, Aasland OG, Babor TF, et al. Development of the alcohol use disorders identification test (AUDIT): WHO Collaborative Project on early detection of persons with harmful alcohol consumption-II. *Addiction*. 1993;88:791–804.
22. NIAAA. *The Physicians Guide to Helping Patients with Alcohol Problems*. NIAAA; 1995.
23. Justice AC, Lasky E, McGinnis KA, et al. Medical disease and alcohol use among veterans with human immunodeficiency infection: A comparison of disease measurement strategies. *Med Care*. 2006;44(8 suppl 2): S52–S60.
24. Freiberg M. Veterans Aging Cohort Study Website. Available at: http://vacohort.org/Images/VACS_data_documentationDCMS.xls. Accessed February 1, 2009.
25. Braithwaite RS, McGinnis KA, Conigliaro J, et al. A temporal and dose-response association between alcohol consumption and medication adherence among veterans in care. *Alcohol Clin Exp Res*. 2005;29: 1190–1197.
26. Naimi TS, Brown DW, Brewer RD, et al. Cardiovascular risk factors and confounders among nondrinking and moderate-drinking U.S. adults. *Am J Prev Med*. 2005;28:369–373.
27. Hadigan C, Meigs JB, Corcoran C, et al. Metabolic abnormalities and cardiovascular disease risk factors in adults with human immunodeficiency virus infection and lipodystrophy. *Clin Infect Dis*. 2001;32: 130–139.
28. Bergersen BM, Sandvik L, Bruun JN, et al. Elevated Framingham risk score in HIV-positive patients on highly active antiretroviral therapy: results from a Norwegian study of 721 subjects. *Eur J Clin Microbiol Infect Dis*. 2004;23:625–630.
29. Currier JS, Taylor A, Boyd F, et al. Coronary heart disease in HIV infected individuals. *J Acquir Immune Defic Syndr*. 2003;33:506–512.
30. Mary-Krause M, Cotte L, Simon A, et al. Increased risk of myocardial infarction with duration of protease inhibitor therapy in HIV infected men. *AIDS*. 2003;17:2479–2486.
31. Klein D, Hurley A, Quesenberry CP Jr, et al. Hospitalizations for CHD and MI among Northern California HIV+ and HIV- men: changes in practive and Framingham Risk scores. Presented at: Conference on Retroviruses and Opportunistic Infections; February 5–9, 2006; Denver, CO.
32. Gardner LI, Klein RS, Szczech LA, et al. Rates and risk factors for condition-specific hospitalizations in HIV infected and uninfected women. *J Acquir Immune Defic Syndr*. 2003;34:320–330.
33. Go AS, Chertow GM, Fan D, et al. Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization. *N Engl J Med*. 2004; 351:1296–1305.
34. Liang TJ, Rehermann B, Seeff LB, et al. Pathogenesis, natural history, treatment, and prevention of hepatitis C. *Ann Intern Med*. 2000;132: 296–305.

New-onset atrial fibrillation and warfarin initiation: High risk periods and implications for new antithrombotic drugs

David A. Garcia[1]; Renato D. Lopes[2]; Elaine M. Hylek[3]

[1]Department of Internal Medicine, University of New Mexico Health Sciences Center, Albuquerque, New Mexico, USA; [2]Duke Clinical Research Institute, Duke University Medical Center, Durham, North Carolina, USA; [3] Department of Medicine, Research Unit-Section of General Internal Medicine, Boston University School of Medicine, Boston, Massachusetts, USA

Summary

Atrial fibrillation is a common condition that increases the risk of stroke in many patients. Although warfarin has been shown to reduce the risk of stroke, many patients who might benefit from anticoagulation do not receive this therapy. Fear of bleeding is the most often cited reason. Several new anticoagulant medications are being studied to determine their efficacy and safety relative to warfarin. Unlike earlier trials that established the superiority of warfarin over placebo, recent trials in atrial fibrillation have enrolled a disproportionate number of patients already taking warfarin. This review suggests that the risk of both haemorrhage and stroke are highest when atrial fibrillation is newly diagnosed and during the initiation of anticoagulant medication. Randomised controlled trials designed to evaluate the safety and efficacy of new antithrombotic agents should include substantial numbers of patients without prior exposure to anticoagulation since these individuals are at the highest risk for bleeding and thromboembolism.

Keywords

Clinical trials, oral anticoagulants, thrombosis, stroke prevention, heart

Correspondence to:
David A. Garcia, MD
Department of Internal Medicine
University of New Mexico Health Sciences Center
University of New Mexico, MSC10 5550
Albuquerque, NM 87131, USA
Tel. : +1 505 925 0404, Fax: +1 505 925 0408
E-mail: davgarcia@salud.unm.edu

Received: July 28, 2010
Accepted after minor revision: September 3, 2010
Prepublished online: September 30, 2010
doi:10.1160/TH10-07-0491
Thromb Haemost 2010; 104: 1099–1105

Introduction

Atrial fibrillation (AF) is a common dysrhythmia; its prevalence is increasing as the population ages (1). Recent projections indicate that by the year 2020, 7.5 million people in the United States will live with AF (2). Because AF results in disorganised electromechanical activity that promotes intra-atrial thrombus formation, the risk of embolic stroke is significant for many patients. The attributable risk of stroke from AF increases significantly with age, from 1.5% for individuals aged 50 to 59 years to 23.5% for individuals aged 80 to 89 years (3). Cardioembolic strokes are associated with substantial morbidity and mortality (4–8).

Between 1989 and 1993, the efficacy of warfarin for stroke prevention in AF was demonstrated decisively by five randomised controlled trials (9–13). Compared to placebo, warfarin use was associated with a 68% relative risk reduction of stroke (14). Despite this dramatic benefit, numerous studies have documented that warfarin is prescribed to only about 50% of at-risk patients with AF (15–18). Older age and perceived bleeding risk are often cited as reasons for not prescribing warfarin. In addition, the narrow therapeutic index of warfarin (combined with its variable dose-response) mandates frequent monitoring which, for many patients, precludes its use (19, 20). Newer antithrombotic agents that match the efficacy of warfarin while also offering a wider therapeutic index would likely increase the use of stroke prevention medications among patients with AF.

The first such new agent to be evaluated was ximelagatran, an oral direct thrombin inhibitor. This drug was studied in two clinical trials enrolling an unprecedented 7,329 patients with AF (SPORTIF-Stroke Prevention using an ORal Thrombin Inhibitor in atrial fibrillation). Based on the previous, placebo-controlled trials conducted in the late 1980s, a primary event (stroke or systemic embolism) rate of 3.1% per year was projected for the patients assigned to receive warfarin (21, 22). Although these studies included elderly populations with a high prevalence of stroke risk factors (▶Table 1), the annual rate of all strokes (ischaemic and haemorrhagic) among patients randomised to warfarin was 2.3% (SPORTIF III) and 1.2% (SPORTIF V) , both lower than the pre-trial estimates. This unexpectedly low rate of thrombotic events among AF patients randomised to warfarin was also reported in a large trial comparing warfarin to dual antiplatelet therapy. In the Atrial Fibrillation Clopidogrel Trial with Irbesartan for prevention of Vascular Events (ACTIVE W) trial, among the 3,371 patients assigned to oral anticoagulation therapy, the rate of stroke (ischaemic or haemorrhagic) plus systemic embolus was 1.5%, also lower than initial projections (23). Similarly, the Birmingham Atrial Fibrillation Treatment of the Aged study (BAFTA) reported low annual stroke rates in both treatment arms (1.8% for warfarin and 3.8% for aspirin) (25).

	AFI (14)	SPORTIF III (21)	SPORTIF V (22)	ACTIVE W (23)	RE-LY (24)
Age, mean	69	70	72	70	72
Gender, male, %	75	70	69	66	63
Heart failure, %	20	34	40	31	32
Hypertension, %	45	72	81	82	79
Diabetes mellitus, %	13	22	25	21	23
Prior stroke/TIA, %	6	24	18	15	20
VKA at entry*, %	10	73	85	78	49

AFI, Atrial Fibrillation Investigators; SPORTIF, Stroke Prevention using an ORal Thrombin Inhibitor in atrial Fibrillation; ACTIVE, Atrial Fibrillation Clopidogrel Trial with Irbesartan for Prevention of Vascular Events; RE-LY, Randomized Evaluation of Long-Term Anticoagulation Therapy; TIA, transient ischaemic attack; VKA, vitamin K antagonist. *Defined differently across trials.

Table 1: Clinical characteristics of AF populations randomised to warfarin.

Although it is possible that the rate of stroke in AF has decreased since the early trials, an alternative hypothesis is that recent trials have enrolled patients who were at lower risk of stroke, despite the similar (or increased) prevalence of stroke risk factors shown in Table 1. Unlike earlier trials designed to establish the superiority of warfarin versus placebo, more recent trials of novel antithrombotic strategies have been designed to establish non-inferiority to warfarin. Thus, the majority of participants in recent trials were taking a vitamin K antagonist (VKA) at baseline. In SPORTIF III and V, 73% and 85%, respectively, of patients randomised to warfarin had been taking a VKA at study entry. The onset of AF was documented to be greater than one year prior to study entry for ≥80% of study participants. Similarly, in the ACTIVE W trial, 77% of participants were taking warfarin at the time of randomisation (warfarin-experienced patients) and duration of AF was six months or greater for 80% of patients and greater than two years for 59% (23). The high proportion of participants who entered these trials already taking warfarin (i.e. "warfarin-experienced") has been an under-appreciated significant difference from earlier AF trials in which greater than 90% of patients were receiving a VKA for the first time (i.e."warfarin-naïve"). A post-hoc analysis of the SPORTIF trial data suggested that the higher proportion of warfarin-naïve participants enrolled in SPORTIF III coupled with greater variability in the INR (International Normalised Ratio) contributed to the higher stroke rate in SPORTIF III versus SPORTIF V (26).

The importance of studying "new users" of a drug has been previously highlighted by Feinstein and Ray (27, 28). The efficacy and safety of new antithrombotic drugs for stroke prevention in AF can be best evaluated among patients who are new to anticoagulant therapy. Consideration of five key elements will demonstrate the chronology bias and survivor bias that impact trials that are dominated by warfarin-experienced patients: 1) bleeding risk during the early phase of anticoagulant therapy, 2) temporal relationship of stroke risk and incident AF, 3) differential early cessation of warfarin therapy among higher risk patients secondary to problems with adherence and tolerability, 4) improved efficacy of warfarin over time resulting from improved control of the INR and decreased INR variability, and 5) potential secondary benefits associated with monthly interface with health care providers (e.g. improved blood pressure control, improved rate control, improved adherence to warfarin and other stroke-modifying therapies), i.e. adherence bias.

I. Major haemorrhage: warfarin-naïve vs. warfarin-experienced

A number of studies have demonstrated that the risk of bleeding on anticoagulant therapy is highest during the period immediately after warfarin is initiated. Landefeld et al. performed a review of medical records from 565 patients prescribed warfarin (for conditions such as AF, venous thromboembolism [VTE] and heart valve replacement) upon hospital discharge. All patients in this cohort were warfarin-naive. The proportion of patients who experienced major bleeding during the first 30 days of outpatient treatment was 3%, 10-fold higher than the calculated monthly risk (0.3%) observed during the subsequent 11 months of follow-up (29). Similarly, in an analysis of data from a randomised controlled trial involving 1,021 patients being started on anticoagulation for VTE, Douketis et al. describe a significant early risk of major haemorrhagic events. Of the 28 major bleeds that occurred during the three months of warfarin treatment in this trial, 21 had occurred within three weeks and 13 had occurred within the first seven days of anticoagulant therapy (30). A meta-analysis of randomised controlled trials or prospective cohort studies of patients being treated for VTE also described a clustering of major bleeding events at the start of anticoagulant therapy: the rate of intracranial haemorrhage (expressed per 100 patient-years) was 5.92 during the first three months of anticoagulant therapy and 0.65 thereafter (31).

A retrospective, multi-center study of 928 consecutive warfarin-treated patients from five anticoagulation clinics yielded similar findings. In order to minimise survivor bias, this analysis included the records not only of patients actively receiving warfarin, but also of patients whose warfarin had been discontinued in the

previous 18–24 months. During a mean duration of follow-up of 1.9 years, being in the early phase of treatment was identified as an independent risk factor for major bleeding. In the first three months of treatment, "serious" bleeding occurred at a rate of 21 episodes per 100 patient-years. The authors found that, compared to the rest of the first year, the second year, and anytime thereafter, the relative risk for serious bleeding during the first three months of treatment was 1.9 (95% confidence interval [CI] 1.3–3.0), 3.0 (95% CI 1.8–4.8), and 5.9 (95% CI 3.8–9.3), respectively (32). When comparing the rate of major haemorrhage during the first 90 days of treatment to the rate of the same outcome after one year of warfarin exposure, a Danish study of warfarin-naïve patients reported a similar increase: incidence rate ratio,1.9 (95% CI 0.8–4.1) (33). ▶Table 2 summarises the evidence that the risk of warfarin-associated bleeding is highest during the initial weeks of treatment.

In an evaluation of all 21,785 patients enrolled in an acute coronary syndrome registry, 2,921 patients were found to have AF (1,700 pre-existing, 1,221 newly diagnosed). Compared to patients in this registry without AF, the patients with AF had increased morbidity and mortality. However, only *new-onset* AF was independently associated with an increased risk of several adverse events, including in-hospital major bleeding (34). Similar results have been demonstrated in a pooled analysis of over 120,000 trial participants with acute coronary syndromes as well as in a group of almost 6,000 patients with ST-segment elevation myocardial infarction treated with primary percutaneous coronary intervention (35, 36). Another large study, the Atrial Fibrillation Follow-up Investigation of Rhythm Management (AFFIRM) trial, has reported similar findings: when compared to patients who entered the trial with a prior history of AF, major haemorrhage was reported to be more common among patients with new-onset AF (37). As noted by Ray, patients initiating therapy are often sicker than prevalent users (28).

This association of major bleeding with *newly diagnosed* (i.e. not pre-existing) AF provides further evidence that a group of patients being exposed to anticoagulation for the first time will have

Table 2: Summary of evidence that the risk of bleeding is highest during the initial days of anticoagulant treatment.

Study	Design + patient population	Length of follow-up	Evidence for "front-loading" of bleeding risk	Comments
Landefeld, 1989 (29)	Retrospective study of patients starting warfarin for a variety of indications	48 months	The monthly risk of major bleeding decreased over time from 3% during the first month to 0.3% per month after the first year of therapy.	All patients identified at the time of hospital discharge
Douketis, 2000 (30)	Analysis of RCT database with 1,021 VTE patients new to warfarin	3 months	28 major bleeds occurred during the initial 3 months of anticoagulation – 21 (75%) of these had occurred with 3 weeks and 13 (46%) had occurred within 7 days of starting anticoagulation	All patients received LMWH "overlap" therapy for the first days of treatment; no AF patients included
Linkins, 2003 (31)	Meta-analysis of 29 RCTs and 4 prospective cohort studies of VTE patients receiving oral anticoagulant therapy for at least 3 months	Variable, but ≥3 months in all cases	Intracranial bleeding during first 3 months of treatment occurred in 1.48% (95% CI 1.40% to 1.56%) of patients. After the first 3 months, the rate of intracranial bleeding was 0.65 (95% CI 0.63 to 0.68) per 100 patient-years.	Definitions of major bleeding differed across studies
Fihn, 1993 (32)	Retrospective study of 5 anticoagulation clinics; 928 patients receiving warfarin for a variety of indications	Median = 1.9 years	Compared to the rest of the first year, the second year, and anytime thereafter, the relative risk for "serious" bleeding during the first 3 months of treatment was 1.9 (95% CI 1.3–3.0), 3.0 (95% CI 1.8–4.8), and 5.9 (95% CI 3.8–9.3) respectively.	Retrospective study, some of the participating centers were not using INR (rather PT ratios) during part of the study
Steffensen, 1997 (33)	Retrospective, single-center cohort study of 682 warfarin-naïve patients being anticoagulated for a variety of reasons	756 treatment-years	The risk of a first major haemorrhagic episode was highest during the first 90 days of treatment; compared with treatment duration above one year, the incidence rate ratio was 1.9.	Although this incidence rate ratio did not reach statistical significance,the numeric difference is consistent with other studies cited
Hylek, 2007 (43)	Prospective, single-center cohort study of 472 warfarin-naïve patients being anticoagulated for AF	12 months	15 of 26 major hemorrhages occurred with 90 days, 11 of 26 within 30 days and 7 withiin the first 2 weeks of warfarin therapy	32% of patients were ≥80 years of age, 33% were identified at hospital discharge, 40% were also taking aspirin

a higher risk of haemorrhage than a group of prevalent users of oral anticoagulation. This may be related to patient-specific, period-specific, or drug-specific factors. Patient-specific factors include the unmasking of subclinical underlying pathologic lesions, labile INR measurements, and concomitant prevalent medications (e.g. anti-platelet therapy) that increase bleeding risk. Period-specific factors confer a transient increase in bleeding risk that would be expected to resolve within a short period of time. For example, the immediate post-hospitalisation period might be characterised by increased use of heparin transition therapy or dietary fluctuation; recently discharged populations will also have a higher prevalence of acute illness (e.g. gastrointestinal mucosal injury, stress-induced gastritis) associated with haemorrhage. Drug-specific factors include half-life, variability in dose-response, tendency toward erratic anticoagulation control with warfarin initiation, and individual rate of INR decay following an episode of excessive anticoagulation. For all of these reasons, it is important that the haemorrhagic risk profile of new anticoagulant drugs with different pharmacokinetic properties be assessed during the period of highest risk, i.e. the 90-day period following therapy initiation.

II. Temporal risk of stroke

Similar to major haemorrhage, the incidence of stroke is also time-dependent. The highest risk of stroke exists at the time of initial presentation with AF. In a registry of 5,477 patients with acute myocardial infarction and left ventricular dysfunction, 1,000 patients had concomitant AF; 655 of these individuals had AF at baseline (pre-existing) and 345 developed AF during the follow-up (median 3.0 years). The adjusted hazard ratio (HR) for 30-day risk of stroke among patients with new-onset AF was 14.6 (95% CI 5.87 – 36.3). In contrast, the adjusted HR for stroke during the whole trial among patients with new-onset AF was 2.29 (95% CI 1.43 – 3.68) (38).

▶Table 3 includes data from a population-based retrospective cohort study where it was noted that the early risk for both haemorrhagic and thromboembolic events was substantially higher than corresponding risks during later time periods of warfarin therapy (39). A third striking example of the high risk for thromboticevents associated with the first 30 days after the diagnosis of AF is seen in an observational study of the Framingham cohort. The authors studied only patients who were not treated with warfarin. Although their principal aim was to derive a stroke risk classification scheme for patients with AF, Wang et al. reported that out of 1,216 patients aged 55 to 94 years, 153 experienced stroke, transient ischaemic attack (TIA) or death within 30 days of AF diagnosis. These 153 individuals, along with others, were excluded from the final analysis of 705 patients who were not treated with warfarin and for whom longitudinal follow-up (mean 4.3 years) was available. The rate of the combined endpoint among the 705 patients who survived the first 30 days following AF diagnosis without suffering stroke, death or TIA was 13.4 per 100 person-years, more than 10 times lower than the corresponding calculated crude incidence rate (153 events per 100 patient-years) among patients who experienced stroke, death or TIA *less than* 30 days after being diagnosed with AF (40).

A study of patients with newly diagnosed AF found that 48% of them had evidence of intra-atrial spontaneous echo contrast by transesophageal echocardiography and 10% had visible intra-atrial thrombus (41). Serial studies have documented complete or near-complete resolution of these thrombi following anticoagulation (42). These observations further highlight the difference between patients with newly diagnosed AF and prevalent warfarin users, among whom visible left atrial thrombus would be very uncommon. The ability of new antithrombotic drugs to antagonise thrombus initiation, propagation, and embolisation will be most rigorously assessed among patients with newly diagnosed AF during this period of heightened risk for thromboembolism.

III. Early attrition of higher risk patients

There is little published data on the proportion of patients in clinical practice who remain on oral anticoagulant therapy over time. If patients with higher stroke risk disproportionately stop oral anticoagulant therapy, then the remaining population of warfarin-experienced patients will be at lower risk of stroke. Furthermore, because patients with a propensity to bleed will most likely experience this adverse event within the first few months of anticoagu-

Event type	Time interval (days)	Person-years at risk	No. of events	No. of censors	Hazard per year of risk
Major bleed	0 – 30	20.6	4	15	0.187
	31 – 90	35.4	4	57	0.116
	91 – 365	68.8	3	113	0.033
	1 – 4.2 years	87.3	7	58	0.068
Any embolism	0 – 30	20.8	7	12	0.328
	31 – 90	35.3	5	62	0.146
	91 – 365	66.7	3	110	0.034
	1 – 4.2 years	84.8	3	59	0.030

Table 3: Risk of haemorrhage or thromboembolism stratified by duration of anticoagulation with warfarin. These data come from a population-based retrospective cohort study that included all residents of Rochester, Minnesota for whom a course of warfarin therapy intended to last for more than four weeks was initiated between September 1, 1987 and December 31, 1989 (39).

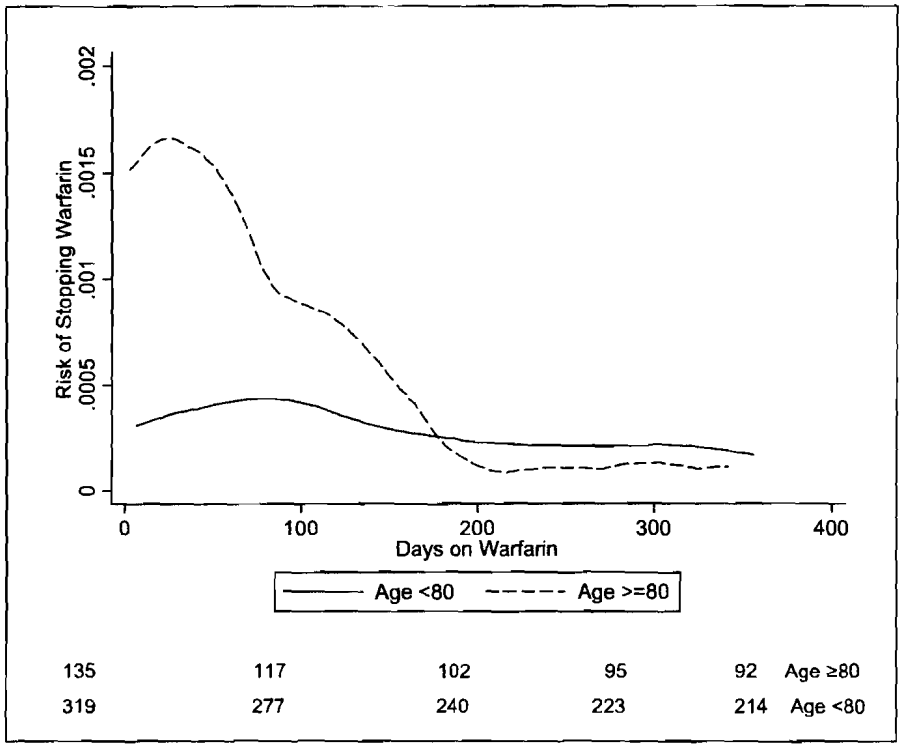


Figure 1: Risk of stopping warfarin in the first year due to perceived safety concerns by age (43). Numbers below graph are the number of patients taking warfarin at that time point. P<0.001, log-rank test.

lant therapy; many such patients will have their warfarin discontinued. Thus, the patients remaining on warfarin will also be at lower risk for haemorrhage. The potential for survivor bias that is introduced by enrolling predominantly prevalent users of warfarin is illustrated by two examples. First, in a cohort study of 472 patients with AF newly starting warfarin, both haemorrhagic events and unplanned discontinuations of therapy occurred more frequently among patients at highest risk of stroke. The risk of stopping warfarin therapy peaked early among patients aged 80 years and older and was similar to that of younger patients at six months (43) (▶Fig. 1). Another study of Medicare beneficiaries tracked adherence to warfarin and INR monitoring post-hospital discharge. The authors found that a subgroup of patients, black and Hispanic individuals, at high baseline risk for stroke were more often lost to follow-up within the first 90 days than their white counterparts (whose baseline stroke risk was lower) (19). Not surprisingly, the black and Hispanic patients in this cohort ultimately suffered higher annual stroke rates (11% and 12% vs. 5% among white patients). Taken together, these studies further emphasize the lower risk profile of patients who manage to remain on warfarin for extended periods of time.

IV. The anticoagulant effect of VKAs stabilises over time

Stability of anticoagulant effect is often not achieved for several months after initiating a VKA. Fluctuations are influenced by diet, drug interactions, genetic variation, and adherence. In a study of 2,223 patients with non-valvular AF, 52% of the INR values were outside the target range (INR 2.0 to 3.0) in the first month, whereas only 30% of INR measurements were not therapeutic after two years of monitoring time (44). Similarly, among 600 adults with AF randomly sampled from three health plans, patients newly started on warfarin at the time of referral to a dedicated anticoagulation clinic tended to have poorer control than patients already taking warfarin (adjusted odds ratio [OR] 0.59; 95% CI 0.35 to 1.08) (45). A comparable association of better anticoagulant control with increased time on warfarin was seen in a study of 254 subjects taking one of three fixed doses of ximelagatran or dose-adjusted warfarin. At study entry, 61% were warfarin-experienced. Among the patients assigned to warfarin, attainment of optimal INR increased from 34% at the start of therapy to 57% at 12 weeks (46).

Enrolment of warfarin-experienced patients in randomized trials confers an advantage to the group assigned to receive warfarin. Because these patients have taken warfarin for an extended period of time, they have become familiar with their own triggers for poor INR control. This training effect results from monthly INR measurements recommended for all warfarin-treated patients. In the ACTIVE W trial of warfarin versus dual antiplatelet therapy, warfarin-naïve patients had less time in the 2.0 to 3.0 range when compared to those who entered the trial on warfarin (warfain experienced), 60.4% versus 64.8% (p<0.001). Warfarin-naïve patients also spent more time with a sub-therapeutic INR, 24.6% versus 19.2% (p<0.001). At three months of follow-up, time in the therapeutic range for the warfarin-naïve group was 57.2% versus 62.4% for the warfarin-experienced group. Not surprisingly, the rates of both vascular events and major haemorrhage were higher in the warfarin-naïve versus warfarin-experienced

group. Patients who entered the trial already on warfarin were also less likely to discontinue the drug, 8.7% versus 15.3% at one year (23). White et al. studied the relationship between INR control and outcomes among 3,587 patients randomised to warfarin in the trials comparing ximelagatran to warfarin. The poor control group, defined as having less than 60% of time with a therapeutic INR, experienced higher rates of major bleeding, thromboembolic events, and annual mortality. Among the reference "good" control group (defined as spending greater than 75% of time with an INR between 2.0 and 3.0), 85% of the individuals entered the study taking warfarin. In contrast, the proportion of warfarin-experienced patients in the corresponding poor (70.8%) and moderate (81.9%) control groups was lower ($p<0.001$) (47). Again, these findings suggest that patients who have taken warfarin for at least several months will spend more time with a therapeutic INR than patients who are new to warfarin.

V. Secondary benefits of frequent medical contact

In addition to the adherence bias introduced by warfarin-experienced patients, long-term use of a medication may be a marker of a more adherent patient population overall. Adherence to antihypertensive medication, lifestyle measures, lipid-lowering agents and diabetes treatment regimens may all collectively decrease an individual's risk of stroke and other thrombotic events over time. The monthly interface with health care providers that occurs with each INR measurement would fortify efforts to control blood pressure, facilitate laboratory monitoring (e.g. measurement of glycated haemoglobin or lipids), minimise patient confusion over treatment plans, and reinforce treatment compliance.

Practical challenges to enrolling warfarin-naïve patients

Enrolling a high proportion of warfarin-naïve patients to participate in a clinical trial is challenging for several reasons. First, patients who are newly starting anticoagulant therapy will be far fewer in number compared to the prevalent user pool. This relative paucity of potential candidates is important because, unlike early placebo-controlled trials that enrolled approximately 250 patients in each arm, trials of non-inferiority require thousands of patients. Warfarin-naïve patients are also more likely to have acute medical conditions that might reduce the likelihood they would be approached about participating in a clinical study. Finally, for patients already taking warfarin, there is no "time limit" during which they need to be randomised. In contrast, the enrolment of patients with newly diagnosed AF is complicated by the fact that they can be considered warfarin-naïve only for a finite period. Given the time-intensive efforts necessary for patient identification, screening, consent, and randomisation, real barriers exist to enrolment of a true inception cohort of patients within a trial setting.

Conclusions

Enrolment of predominantly warfarin-experienced patients in clinical trials results in under-ascertainment of early events associated with initiation of anticoagulant therapy and newly diagnosed AF. The "warfarin surviving" patients also may introduce an adherence bias because of improved INR and blood pressure control over time. Before definitive conclusions about the relative safety and efficacy of novel anticoagulant drugs can be drawn, it is important that each of these agents be studied in warfarin-naïve patients because these patients are at the highest risk for both stroke and major haemorrhage.

Conflicts of interest

Dr. Garcia has served in an advisory capacity for Boehringer Ingelheim, Bristol-Myers Squibb, and Ortho McNeill Jansen. Dr. Lopes has received research funding from Bristol-Myers Squibb. Dr. Hylek has served in an advisory capacity for Astellas, Bayer, Boehringer-Ingelheim, Bristol-Myers Squibb, Genentech, Merck, Medtronic, Pfizer, and Sanofi-Aventis and received research funding from Bayer Healthcare and Bristol-Myers Squibb.

References

1. Go AS, Hylek EM, Phillips KA, et al. Prevalence of diagnosed atrial fibrillation in adults: National implications for rhythm management and stroke prevention: The anticoagulation and risk factors in atrial fibrillation (atria) study. J Am Med Assoc 2001; 285: 2370–2375.
2. 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: 119–125.
3. Wolf PA, Abbott RD, Kannel WB. Atrial fibrillation as an independent risk factor for stroke: The framingham study. Stroke 1991; 22: 983–988.
4. Hart RG, Pearce LA, Miller VT, et al. Cardioembolic vs. Noncardioembolic strokes in atrial fibrillation: Frequency and effect of antithrombotic agents in the stroke prevention in atrial fibrillation studies. Cerebrovasc Dis 2000; 10: 39–43.
5. Hylek EM, Go AS, Chang Y, et al. Effect of intensity of oral anticoagulation on stroke severity and mortality in atrial fibrillation. N Engl J Med 2003; 349: 1019–1026.
6. Longstreth WT, Jr., Bernick C, Fitzpatrick A, et al. Frequency and predictors of stroke death in 5,888 participants in the cardiovascular health study. Neurology 2001; 56: 368–375.
7. Miller VT, Pearce LA, Feinberg WM, et al. Differential effect of aspirin versus warfarin on clinical stroke types in patients with atrial fibrillation. Stroke prevention in atrial fibrillation investigators. Neurology 1996; 46: 238–240.
8. Sherman DG, Goldman L, Whiting RB, et al. Thromboembolism in patients with atrial fibrillation. Arch Neurol 1984; 41: 708–710.
9. Stroke prevention in atrial fibrillation study. Final results. Circulation 1991; 84: 527–539.
10. Ezekowitz MD, Bridgers SL, James KE, et al. 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.
11. Petersen P, Boysen G, Godtfredsen J, et al. 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.

12. The Boston Area Anticoagulation Trial for Atrial Fibrillation Investigators. The effect of low-dose warfarin on the risk of stroke in patients with nonrheumatic atrial fibrillation. N Engl J Med 1990; 323: 1505–1511.
13. EAFT (European Atrial Fibrillation Trial) study group. Secondary prevention in non-rheumatic atrial fibrillation after transient ischaemic attack or minor stroke. Lancet 1993; 342: 1255–1262.
14. Risk factors for stroke and efficacy of antithrombotic therapy in atrial fibrillation. Analysis of pooled data from five randomized controlled trials. Arch Intern Med 1994; 154: 1449–1457.
15. Waldo AL, Becker RC, Tapson VF, et al. 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.
16. Cohen N, Almoznino-Sarafian D, Alon I, et al. Warfarin for stroke prevention still underused in atrial fibrillation: Patterns of omission. Stroke 2000; 31: 1217–1222.
17. Gage BF, Boechler M, Doggette AL, et al. Adverse outcomes and predictors of underuse of antithrombotic therapy in medicare beneficiaries with chronic atrial fibrillation. Stroke 2000; 31: 822–827.
18. Hylek EM, D'Antonio J, Evans-Molina C, et al. Translating the results of randomized trials into clinical practice: The challenge of warfarin candidacy among hospitalized elderly patients with atrial fibrillation. Stroke 2006; 37: 1075–1080.
19. Birman-Deych E, Radford MJ, Nilasena DS, et al. Use and effectiveness of warfarin in medicare beneficiaries with atrial fibrillation. Stroke 2006; 37: 1070–1074.
20. Johnston JA, Cluxton RJ, Jr., Heaton PC, et al. Predictors of warfarin use among ohio medicaid patients with new-onset nonvalvular atrial fibrillation. Arch Intern Med 2003; 163: 1705–1710.
21. Executive Steering Committee for the SPORTIF III Investigators. Stroke Prevention with the oral direct thrombin inhibitor ximelagatran compared with warfarin in patients with non-valvular atrial fibrillation. Lancet 2003; 362: 1691–1698.
22. Albers GW, Diener HC, Frison L, et al. Ximelagatran vs warfarin for stroke prevention in patients with nonvalvular atrial fibrillation: A randomized trial. J Am Med Assoc 2005; 293: 690–698.
23. Connolly S, Pogue J, Hart R, 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: 1903–1912.
24. Connolly SJ, Ezekowitz MD, Yusuf S, et al., and the RE-LY Steering Committee and Investigators. Dabigatran versus Warfarin in Patients with Atrial Fibrillation. N Engl J Med 2009; 361: 1139–1151.
25. Mant J, Hobbs FD, Fletcher K, et al. 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.
26. Hylek EM, Frison L, Cupples LA. Disparate Stroke Rates on Warfarin Among Contemporaneous Cohorts with Atrial Fibrillation: Potential Insights into Risk from a Comparative Analysis of SPORTIF III versus SPORTIF V (Stroke Prevention using an ORal Thrombin Inhibitor in atrial Fibrillation) Stroke 2008; 39: 3009–3014.
27. Feinstein AR. Clinical biostatistics. Xi. Sources of 'chronology bias' in cohort statistics. Clin Pharmacol Ther 1971; 12: 864–879.
28. Ray WA. Evaluating medication effects outside of clinical trials: New-user designs. Am J Epidemiol 2003; 158: 915–920.
29. Landefeld CS, Goldman L. Major bleeding in outpatients treated with warfarin: Incidence and prediction by factors known at the start of outpatient therapy. Am J Med 1989; 87: 144–152.
30. Douketis JD, Foster GA, Crowther MA, et al. Clinical risk factors and timing of recurrent venous thromboembolism during the initial 3 months of anticoagulant therapy. Arch Intern Med 2000; 160: 3431–3436.
31. 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.
32. Fihn SD, McDonell M, Martin D, et al. Risk factors for complications of chronic anticoagulation. A multicenter study. Warfarin optimized outpatient follow-up study group. Ann Intern Med 1993; 118: 511–520.
33. Steffensen FH, Kristensen K, Ejlersen E, et al. Major haemorrhagic complications during oral anticoagulant therapy in a danish population-based cohort. J Intern Med 1997; 242: 497–503.
34. Mehta RH, Dabbous OH, Granger CB, et al. Comparison of outcomes of patients with acute coronary syndromes with and without atrial fibrillation. Am J Cardiol 2003; 92: 1031–1036.
35. Lopes RD, Pieper KS, Horton JR, et al. Short- and long-term outcomes following atrial fibrillation in patients with acute coronary syndromes with or without ST-segment elevation. Heart 2008; 94: 867–873.
36. Lopes RD, Elliott LE, White HD, et al. Antithrombotic therapy and outcomes of patients with atrial fibrillation following primary percutaneous coronary intervention: results from the APEX-AMI trial. Eur Heart J 2009; 30: 2019–2028.
37. DiMarco JP, Flaker G, Waldo AL, et al. Factors affecting bleeding risk during anticoagulant therapy in patients with atrial fibrillation: Observations from the atrial fibrillation follow-up investigation of rhythm management (affirm) study. Am Heart J 2005; 149: 650–656.
38. Lehto M, Snapinn S, Dickstein K, et al. Prognostic risk of atrial fibrillation in acute myocardial infarction complicated by left ventricular dysfunction: The optimal experience. Eur Heart J 2005; 26: 350–356.
39. Gitter MJ, Jaeger TM, Petterson TM, et al. Bleeding and thromboembolism during anticoagulant therapy: A population-based study in rochester, minnesota. Mayo Clinic Proc 1995; 70: 725–733.
40. Wang TJ, Massaro JM, Levy D, et al. A risk score for predicting stroke or death in individuals with new-onset atrial fibrillation in the community: The framingham heart study. J Am Med Assoc 2003; 290: 1049–1056.
41. Manning WJ, Silverman DI, Waksmonski CA, et al. Prevalence of residual left atrial thrombi among patients with acute thromboembolism and newly recognized atrial fibrillation. Arch Intern Med 1995; 155: 2193–2198.
42. Tsai LM, Chen JH, Lin LJ, et al. Natural history of left atrial spontaneous echo contrast in nonrheumatic atrial fibrillation. Am J Cardiol 1997; 80: 897–900.
43. 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.
44. Jones M, McEwan P, Morgan CL, et al. 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.
45. Menzin J, Boulanger L, Hauch O, et al. Quality of anticoagulation control and costs of monitoring warfarin therapy among patients with atrial fibrillation in clinic settings: A multi-site managed-care study. Ann Pharmacother 2005; 39: 446–451.
46. Petersen P, Grind M, Adler J. Ximelagatran versus warfarin for stroke prevention in patients with nonvalvular atrial fibrillation. Sportif ii: A dose-guiding, tolerability, and safety study. J Am Coll Cardiol 2003; 41: 1445–1451.
47. 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: 239–245.

JOURNAL OF WOMEN'S HEALTH
Volume 19, Number 4, 2010

DOI: 10.1089/jwh.2009.1627

Progress and Priorities in the Health of Women and Girls: A Decade of Advances and Challenges

Francisco A.R. Garcia, M.D., M.P.H.,[1] Karen M. Freund, M.D., M.P.H.,[2] Michelle Berlin, M.D., M.P.H.,[3] Kathleen B. Digre, M.D.,[4] Donald J. Dudley, M.D.,[5] Rose S. Fife, M.D., M.P.H.,[6] Geralde Gabeau, M.S.M., M.P.H.,[2] Stacie E. Geller, Ph.D.,[7] Jeanette H. Magnus, M.D., Ph.D.,[8] Justina A. Trott, M.D.,[9] and Hilary F. White, B.S.[2]

Abstract

Objective: Following the initial wave of federal support to address women's health, there is a need to assess successes and determine the next priorities to advance the health of women. The objective of this study was to systematically collect expert opinion on the major advances in women's health in the past decade and priorities for women's health research and service in the coming decade.
Methods: We utilized a Delphi method to query the leadership from academic and community Centers of Excellence in Women's Health, as designated by the Department of Health and Human Services. Leaders from 36 of the 48 centers responded to a series of questions about the major advances and critical indicators to evaluate future needs in women's health. We utilized a social ecology model framework to organize the responses to each question.
Results: The experts identified increased health education for women and increased empowerment of women across multiple spheres as the major advances positively impacting the health of women. The experts selected the following areas as the most important indicators to measure the status of the health of women in the future: health education and promotion, rates and impact of interpersonal violence against women, and access to healthcare. The major advances and measures of the health of women did not focus on specific changes to individual women in illness management, clinical care, or individual behavioral change.
Conclusions: As we move to address health reform, we must be able to recognize and incorporate a broad perspective on public health and policy initiatives critical to the health and wellness of women and girls and, therefore, central to the well-being of the nation.

Introduction

Focus on all aspects of women's health has increased dramatically in the last two decades. The recognition that women were less likely to be included in clinical trials led to policy changes, including new National Institutes of Health (NIH) Requirements for the Inclusion of Women and Minorities as Subjects in Clinical Research.[1] Landmark studies that examined female-male differences provided important insights and unexpected findings that informed clinical and public health practices and led to a variety of federal initiatives to address the unique health and wellness needs of women and girls. These efforts resulted in the congressional mandate that established the Office for Research in Women's

[1]Obstetrics and Gynecology and Public Health, Center of Excellence in Women's Health, The University of Arizona, Tucson, Arizona.
[2]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, Massachusetts.
[3]Department of Obstetrics and Gynecology, and Center for Women's Health, Oregon Health and Science University, Portland, Oregon.
[4]Department of Neurology, Ophthalmology, Obstetrics and Gynecology, University of Utah, Salt Lake City, Utah.
[5]Department of Obstetrics and Gynecology, University of Texas Health Science Center at San Antonio, San Antonio, Texas.
[6]Indiana Clinical and Translational Sciences, Indiana University School of Medicine, Indianapolis, Indiana.
[7]Department of Obstetrics and Gynecology, Center for Research on Women and Gender, University of Illinois, Chicago College of Medicine, Chicago, Illinois.
[8]Maternal and Child Health Section, Department of Community Health Sciences, Tulane University School of Public Health and Tropical Medicine, New Orleans, Louisiana.
[9]Department of Internal Medicine, Women's Health Services, University of New Mexico School of Medicine, Albuquerque, New Mexico.

Health (ORWH) at the NIH, and the establishment of the Offices on Women's Health (OWH) at the Department of Health and Human Services (DHHS) and the Office of Women's Health at the Food and Drug Administration (FDA).

A subsequent initiative was the establishment and funding of academic and community national Centers of Excellence in Women's Health (CoEs) by the DHHS Office on Women's Health from 1995 to 2007. These interdisciplinary model programs were funded to develop comprehensive, coordinated, and integrated female-focused programs that encompassed clinical care, gender-based research, professional education, community outreach, leadership development, and mentoring.

These comprehensive women's health programs have a lifespan perspective and address the needs of women and girls within the bio-psycho-social continuum. This view presents a challenge to a traditional medical model when one aims to develop a "roadmap" for women's health, particularly since unifying measures or benchmarks of success and accomplishments are clearly needed in women's health.

It is critical to view each woman as an individual, living in and affected by a specific physical and social environment: her community. Within that environment, many factors impinge upon her health from the moment of conception until her death. In an effort to evaluate the impact of both the individual and the environment, we have reached beyond the medical model to other fields. Recognizing the multiple influences that affect the health of women, we sought to understand the current status of the health of women and project emerging needs as we move forward. We used the multi-level Bronfenbrenner's Ecology of Human Development[2] or Social Ecology Model to survey opinion leaders in the CoE programs about their understanding of women's health today, from the individual, interpersonal, community, institutional/organizational, and public policy levels. The model recognizes that decisions and behavior are the result of a synergy among these levels and cannot be interpreted well without a good understanding of the context in which such decisions and behaviors occur.[2]

Materials and Methods

The goals of this project were to describe the most significant advances in women's health during the past decade and to identify important indicators that assess whether improvements are being achieved. This was accomplished by performing a theory-based survey of directors or senior leaders of the CoE programs funded by the DHHS OWH, as of 2006.[3] These centers, and the populations they served, broadly reflected the geographic and racial/ethnic diversity of the nation. They included: (1) academically based Centers of Excellence in Women's Health; (2) community-based Centers of Excellence (CCoEs); (3) Rural/Frontier Women's Health Coordinating Centers (RFCC); and (4) Region VIII Demonstration Sites in the intermountain West and Dakota regions.

All programs, hereafter referred to as CoEs, funded by OWH in 2006 were invited to participate in the survey. We collected the opinions of 36 of the 48 designated centers using a Delphi process. The Delphi process is a method for achieving consensus by completion of two or more rounds of questionnaires, with feedback on the group's and one's own results between administrations.[4] The process was originally developed for participants who could not meet; however, group discussion between rounds can, and were, used to facilitate consensus.[4] Even when consensus is not achieved, the method allows one to assess the degree of agreement in a group. For this study we used a three-round, conventional Delphi method.

We were interested in two specific questions: (1) What have been the most important advances in women's health in the past decade? and (2) What are the most important indicators to observe in the future to evaluate whether the health of women is improving?

The flow of the Delphi process is presented in Figure 1. In the first round, the writing group developed the strategy for organizing responses using the social ecology model and submitted a list of possible responses to both of the questions. The results of the first round were submitted in written form to all CoE directors to review at their sites for the second round of the Delphi process. Directors were given the option of reviewing individually or obtaining the consensus from the leadership at their site. They were asked to expand and list additional responses to each question not included in the list. They were then asked to provide written responses to the two principal questions. The third round of the Delphi process took place at an in-person conference of all 48 directors or their designees. The written compilation of responses from the second round of the Delphi process was provided to all directors for review. The directors were then divided into five working groups each, which formed their own summary of the key responses and presented their findings to the entire group. Following this process, each director or designee was given the same written document and asked to list his/her top responses to the two questions.

After the third round of the Delphi process, two authors (FG and KF) reviewed these responses and collapsed the raw responses into major themes. This summary was placed into two tables to demonstrate the interrelatedness of the themes across the five levels of the model. Internal feedback was provided by the writing group to develop the themes. The entire writing group then reviewed and revised the themes into the final format presented.

Results

Two tables summarize the major themes identified within the context of social ecology model structure. Table 1 contains advances in women's health and Table 2 lists indicators to assess progress in women's health. The institutional/organizational and public policy levels were collapsed into one column, to reflect the substantial overlap in the responses for these two categories. The themes (rows) were ordered based on the number of levels (columns) into which each theme fell. Boxes were bolded to indicate that at least 10 CoEs included this idea in their responses. Themes that were both included by fewer than 10 CoEs and were mentioned under only one level (column) were excluded from the tables. The following section summarizes the main patterns and findings derived from these tables.

Advances in women's health

Health education was cited across all levels of the model (Table 1). At the individual level, this included access by patients to health education, community health education

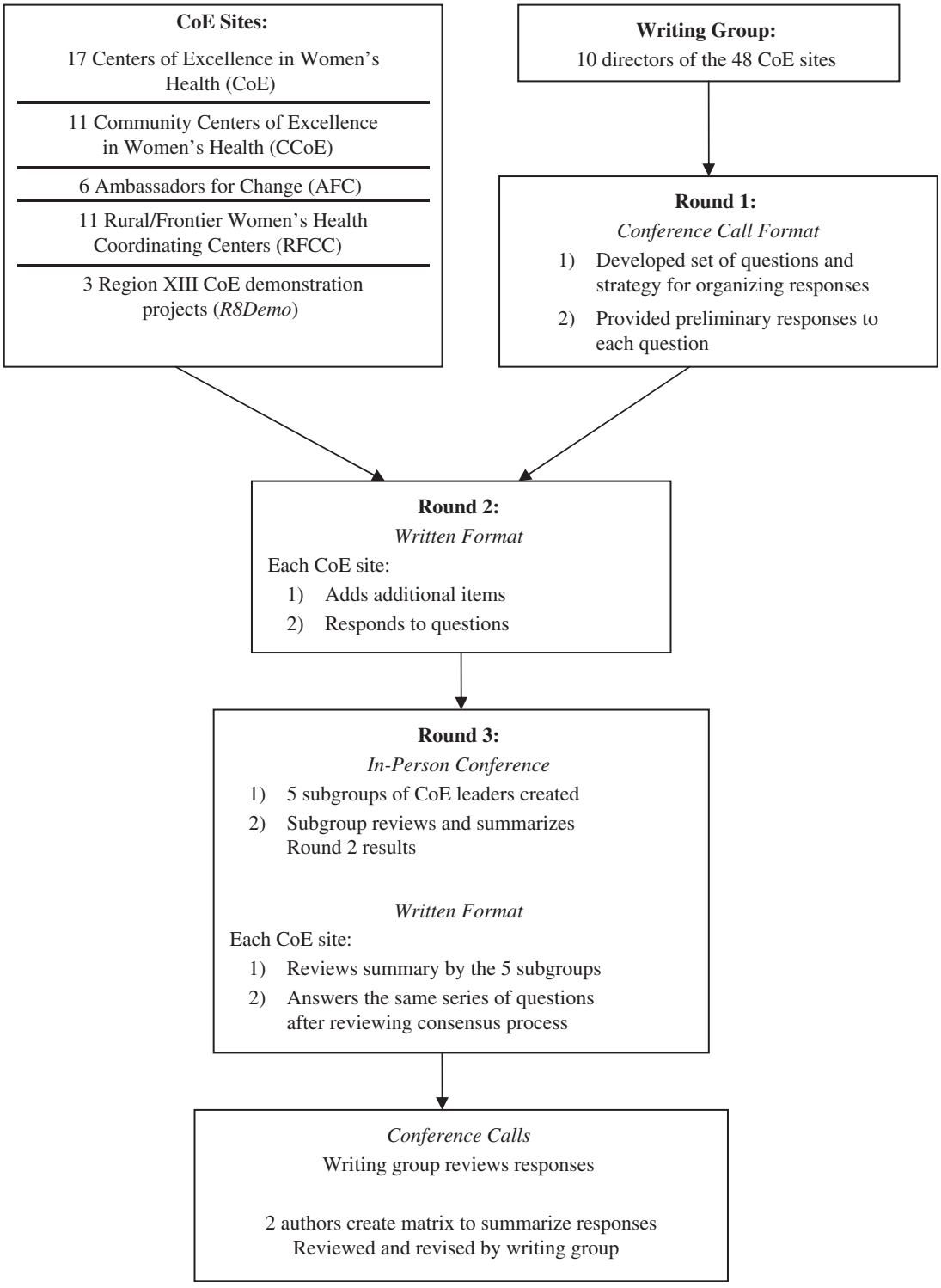


FIG. 1. Flow of the Delphi process.

programs, and increased knowledge of health issues. At the interpersonal level, the ubiquity and availability of accurate, relevant, and contextual health information for the individual woman and her family was cited by a number of respondents. By far the most frequently cited advance in community level women's health was the recognition of community-based partnerships in various forms as important and effective means for promoting health through health education. At institutional/public policy levels, respondents cited workplace-based wellness and health education programs as important advances, as well as increased public awareness of specific women's health issues, such as nonreproductive-related women's health, geriatric issues, disparities, and healthcare access and coverage issues.

Empowerment of women was cited at three levels: individual, interpersonal, and community. The increased

Table 1. Advances in Women's Health

	Individual	*Interpersonal*	*Community*	*Institutional/organizational and public policy*
Health education	Access by patients to health education Community health education programs Increased knowledge of health issues	Availability of accurate, relevant, and contextual health information	Via community-based partnerships	Workplace-based wellness and health education programs
Empowerment of women	Increased awareness of women as consumers of health Active/vocal role in decision-making Attending to self-care Personal initiative Activist role	Improved self-efficacy of women	Via community-based partnerships	
Employment-related factors/equity in the workplace		Recognition of workplace environmental influences on personal health	Worksite wellness programs Flexibility to attend to personal and family issues Economic sustainability	Advances in policies to promote gender equity in the workplace (especially salary equity) Policies in child/eldercare, medical/family leave, flex time, part-time employment
Safety/violence prevention		Child abuse Intimate partner violence Other domestic violence	Safe/low-crime neighborhoods	
Support and communication		Healthy/high quality interpersonal relationships in home and work environment Social support networks/groups (support personal development and self-efficacy)	*See* Community-based partnerships/initiatives	
Environmental factors		Recognize home and workplace environmental influences on personal health	Availability of public spaces for physical activity and recreation Safe and low-crime neighborhoods Adequate housing Tobacco-free communities	

Recognition of women's health issues	Comprehensive approach to health of women Sex and gender differences between men and women			
Community-based partnerships/ initiatives			Community-based organizations Faith-based organizations Neighborhoods Informal and formal social networks Lay health educators	
Women in leadership roles				Initiatives/programs to increase proportion of women leaders Evidence of an increase in women leadership
Focus on women's health				Financial support Focus of training, curriculum, and research on women's health
Women's health/ gender-based research				National initiatives promoting women's health research (e.g., Women's Health Initiative) Research offices within federal agencies (i.e., NIH, DDA, etc.) Specific funding Regulations requiring inclusion of women in clinical trials
Federal initiatives in women's health				Advent of federal and state support for women's health agencies to promote and sustain initiatives in women's health (e.g., BIRCWH, CoEs, national Offices on Women's Health in DHHS, NIH, FDA, etc.)

= ≥10 responses.
= <10 responses.

Table 2. Indicators to Measure Progress in Women's Health

	Individual	*Interpersonal*	*Community*	*Institutional/organizational and public policy*
Health education/ promotion	Access to health education Increases in health literacy Increased self-efficacy or participation in one's own health care	Access, exposure, and quality of health education programs and materials	Improving consumer information to support healthy lifestyle choices Availability of health education to students Quality of education and achievement in general	Employer support for health promotion and education activities
Rates of interpersonal violence	Rates of interpersonal violence	Rates of incidence of and screening for intimate partner violence and domestic violence Perceived safety	Rates of domestic violence, neighborhood crime, and incarceration Perceived safety of communities	Employer-based resources for domestic violence victims
Access to healthcare	Overall access to healthcare Access to a personal care provider and wellness care Mental health care Fertility control and prenatal care Insurance coverage		Availability of community-based and public sector health care providers and insurance	Universal health insurance Access to primary care and screening services Funding of essential therapeutics Access to mental health as part of employer-funded health insurance
Patient satisfaction	Patient satisfaction and quality-of-life indicators (as outcomes of healthcare)	"No show" and compliance rates Perceived empowerment Satisfaction with services provided		Patient satisfaction with their healthcare system
Preventive health	Health screenings Measures of lifestyle or behavior, including exercise, nutrition, and BMI, smoking, and drug abuse			Preconception health Infant wellness Family planning Nutrition Rates of specific screening tests (e.g., mammography)
Outcomes (prevalence of disease and morbidity/ mortality)	Incidence, prevalence, management markers, morbidity and mortality of chronic disease conditions and comparing these by sex and racial/ethnic group for disparities in outcomes			

Social measures of wellness	Economic indicators (poverty levels, rates of housing, etc)			
Relationship factors		Quality and effectiveness of communication with partner, family, and healthcare providers		
Community-level policy			"Grass roots" organizations Neighborhood resource centers Smoking bans Infrastructure to support physical activity Community alliances Lobbying policy makers	
Federal support for women's health research				NIH support for research training programs (BIRCWH) HHS support for CoE/CCOE program
Women in leadership positions				Within the institution

= ≥10 responses.
= <10 responses.

empowerment of women in their own health was the most commonly noted advance at the individual level. This was described as increased awareness of women as consumers of healthcare, women taking a more active and vocal role in decision-making, women attending more to self-care, and personal initiative, and women taking more of an activist role. At the interpersonal level, a number of respondents cited the development of social support networks to improve self-efficacy of women as an important advance. At the community level, respondents identified the development of community-based partnerships and initiatives aimed at empowering women as very important.

Employment-related factors and equity in the workplace were cited at the interpersonal, community, and institutional/public policy levels. At the interpersonal level, respondents noted the recognition of workplace environmental influences on personal health as an important advance. At the community level, employment-related factors such as worksite wellness programs, flexibility to attend to personal and family issues, and economic sustainability were mentioned by a minority of respondents. At the institutional/public policy level, respondents cited advances in policies to promote gender equity in the workplace (especially salary equity), as well as "female-friendly" and "family-friendly" workplace policies related to child and elder care, medical and family leave, flexible hours, and part-time employment, as critical advances.

Safety and violence prevention were cited at the interpersonal and community levels. Programs to promote safety and security of women were the most frequently cited advance in women's health at the interpersonal level. Specifically, respondents mentioned advances in prevention programs and policies of interpersonal violence, including child abuse, intimate partner violence, and other forms of domestic-related violent behaviors. At the community level, initiatives for safe and low-crime neighborhoods were considered to be important advances.

Support and communication were also mentioned at both the interpersonal and community levels, and constituted an especially important theme at the interpersonal level. Specifically, respondents cited healthy/high-quality interpersonal relationships in the home and work environment, as well as social support networks to promote personal development and self-efficacy, as important advances in women's health at the interpersonal level. Support and communication were also important at the community level, and the most frequently cited advance at this level was the recognition of community-based partnerships and initiatives (including community-based organizations, faith-based organizations, neighborhoods, informal and formal social networks, and lay health educators) as crucial mechanisms to support and promote health education, health promotion, and the empowerment of women.

Environment-related factors were another theme cited at both the interpersonal and community levels. At the interpersonal level, these included the recognition of home and workplace environmental influences on personal health. At the community level, advances in environmental factors included the availability of public spaces for physical activity and recreation, safe and low-crime neighborhoods, adequate housing, and promotion of tobacco-free communities.

Twelve themes were mentioned in just one of the four levels. Six of these were cited by more than 10 CoEs (shown in Table 1). At the individual level, CoEs noted the recognition of broader women's health issues, including a comprehensive approach to the health of women and recognition of sex and gender differences between men and women, as an important advance in women's health. At the community level, respondents identified advances regarding community-based partnerships/initiatives. At the institutional/public policy level, multiple CoEs cited women in leadership roles, an increased focus on women's health, federal/national funding for women's health/gender-based research (e.g., the Women's Health Initiative and federal regulations regarding the inclusion of women and minorities in clinical trials), and federal initiatives in women's health (e.g., the NIH Building Interdisciplinary Research Careers in Women's Health[5] and CoE programs[6]) as important advances.

Indicators to measure progress in women's health

The second set of questions asked the group of experts to identify measures or indicators that would best capture the status of the health of women and girls. Health education and health promotion are indicators that were found across all four levels of the model (Table 2). At the individual level, respondents cited access to health education, health literacy, and self-efficacy or participation in one's own healthcare as important indicators. Interpersonal-level indicators regarding health education/promotion included access/exposure and quality of health education programs and materials. At the community level, measures of consumer information to support healthy personal and lifestyle choices, measures of the educational content available to students regarding health-related topics, and quality of education and achievement in general were cited.

Rates of interpersonal violence were mentioned across all four levels. At the individual and interpersonal levels, respondents cited rates of intimate partner violence (domestic violence). At the interpersonal level, rates of screening for interpersonal violence, as well as perceived safety, were noted as important by a number of contributors. Community-level indicators concerning interpersonal violence included rates of domestic violence, neighborhood crime, and female incarceration, as well as perceived safety of communities. At the institutional/public policy level, respondents cited availability of employer-based resources for domestic violence victims as important.

Access to healthcare was cited as a critical measure at the individual, community, and institutional/public policy levels. At the individual and community levels, access to healthcare was the most commonly cited indicator for the measurement of women's health. At the individual level, indicators included overall access to health care, access to a primary care provider (PCP) and wellness care, mental healthcare, fertility control and prenatal care, and insurance coverage. At the community level, respondents cited availability of community-based and public sector healthcare providers and insurance as important measures. Indicators regarding access to healthcare at the institutional/public policy level included universal health insurance, access to primary care and screening coverage, funding for essential therapeutics, and access to mental health care as part of employer-funded health insurance.

Preventive health indicators were cited as important measures of health by a number of contributors at both the individual and institutional/public policy levels. At the individual level, respondents cited health screenings and mea-

sures of lifestyle and behavior, including exercise, nutrition and BMI, smoking, and drug abuse. Preventive health indicators at the institutional/public policy level included the availability of preventive service programs for women and families, including programs for preconception health, infant wellness, family planning, nutrition, and rates of specific screening tests, such as mammography.

Twelve measurement indicator themes were mentioned in just one of the four levels. Six of these were cited by more than 10 CoEs (shown in Table 2). Individual level outcomes of healthcare, including the incidence, prevalence, management markers, and morbidity and mortality of chronic disease conditions, compared by both sex and racial/ethnic group for disparities in outcomes, were emphasized. Social measures of wellness (including economic indicators such as poverty levels, rates of housing, and rates of interpersonal violence), were mentioned as important indicators to measure women's health at the individual level. At the interpersonal level, multiple sites mentioned relational factors, including measures of the quality and effectiveness of interpersonal relationships and communications with partner, family, and healthcare providers. At the community level policy engagement pertaining to "grass-roots" organizations, neighborhood resource centers, smoking bans, infrastructure to support physical activity, influence over school-based policies (e.g., dietary and physical education), formation of community alliances, and lobbying of policy makers were highlighted. At the institutional/public policy level, multiple sites mentioned federal support for women's health services and research (e.g., BIRCWH[5] and CoE/CCOE[6]). Multiple sites also listed the number of women in leadership positions as an important institutional indicator of women's health.

Discussion

We asked a group of experts in the healthcare of women to list the critical measures of the health of women, and the most useful advances in the health of women. Of interest, the major advances and measures of the health of women did not focus on specific changes to the individual women in illness management, clinical care, or individual behavioral change. Rather, they dealt with changes to the environment to promote health, and changes in women's self-empowerment to address their own health and wellness.

These findings are critical as we engage in a discussion about changes to the healthcare system, especially whether improvements in access to care will eliminate health disparities. Critical to the discussion of the nation's health is the assertion that healthcare access, although significant, cannot solve the spectrum of health needs of our population. Addressing health and wellness solely at the level of the individual is insufficient, since individual change alone does not address the spectrum of contextual issues that impact the health and wellness of the population. The social ecology model is informative since it offers a broader perspective for improving the health of the nation and, specifically, the health of women. Although health insurance reform is needed, the health of women and the factors associated with women's health go beyond coverage for illness and disease prevention, and must address the full spectrum of relevant social and public health issues. Furthermore, our concept of the health of any individual must go beyond the notion that the individual is solely responsible for his/her health. While personal responsibility for one's health and wellness is a necessary element, much of what influences health and wellness lies within spheres beyond the individual, including interpersonal, community, institutional, and public policy arenas.

The expert panel assembled to conduct this Delphi process represents an interdisciplinary approach to addressing health and wellness. The advances and the greatest needs in the health of women are therefore viewed across the spectrum of individual, interpersonal, community, and institutional/public policy issues. This approach has direct policy implications for changes to support the wellness of women in our society. The establishment of the Offices on Women's Health in the U.S. Department of Health and Human Services in 1991 facilitated the establishment of the National Centers of Excellence in Women's Health, a program aimed at increasing the focus on women's health across the United States, first in academic health centers and then in community settings.[6] Each CoE independently developed its version of the model based on the culture and policy of its institution, partner organizations, community, and state. However, all CoE models were charged to provide comprehensive, coordinated, and integrated gender-based programs in clinical care, gender-based research, professional education, community outreach, and leadership development/mentoring.[7]

Interpersonal violence is a good example of a health condition or problem of specific relevance to women, for which a social ecology approach is critical to effect change. Our panel identified interpersonal violence as one of the key areas—spanning the individual, interpersonal, community, and institutional/public policy realms—where metrics are needed to measure progress in the wellness of women. Each year, approximately 1,300 deaths and 2 million injuries among women are attributed to intimate partner violence.[8] Interpersonal violence against women results in increased morbidity and significant increases in physical and mental health problems.[9] The significant physical and mental health consequences of interpersonal violence also result in large long-term health expenditures, including an average increase in 3-year healthcare costs of $1,700 per abused woman[10] and an estimated cost of violence of 3.3% of the U.S. gross domestic product.[11] Yet, while individual programs to guarantee women's safety are critical, these interventions do little to address the prevention of injury and morbidity from interpersonal violence. Our analyses highlight a number of critical measures with which to assess improvements for the prevention of violence to women. These include increasing the rates of women reporting active threats or injury, programs to proactively screen and address safety issues in the clinical setting and to address safety issues on a community basis (including enforcement of restraining orders and programs to ensure safety overall in neighborhoods), and institutional programs to ensure the safety of individuals in the workplace and other institutional settings.

A number of markers of improvements in the health of women fall outside the realm of medicine. Pay equity and training and mentorship of women into leadership positions are examples of important indicators that lie in the institutional and public policy domains. Such nonmedical factors are just as crucial to the health of women as are medical concerns. These issues reflect the larger social relationship between a nation and its citizens and illustrate how the context of

women's lives has profound implications for the national economy, workplace productivity, education, the environment, and, of course, public health.

The strength of this paper is in the wide geographic and interdisciplinary natures of the institutions and experts who responded to the process. One limitation of the model, however, is that it views women as passive within the process and therefore being affected by their environment. In reality, women actively impact their environment and community, an aspect that was not captured in our model.

Conclusions

Healthcare reform has re-entered the national dialogue, providing an important opportunity to consider how best to impact the health and wellness of America's women and girls within the larger context of the nation's health. The Centers of Excellence provide a model of implementing healthcare reform through a broad array of programs and outreach mechanisms. The advances in women's health identified in this study reflect areas of relative success in communities and settings where a well-developed multidisciplinary model has been applied. Thematic areas (e.g., patient-centered care/customer satisfaction) and levels of intervention (community) that have received relatively less attention were identified and should be considered for future funding and support. Furthermore, our analysis identifies concrete and meaningful metrics to track as new programs and strategies develop to address the health and wellness of our population and of women and girls, in particular.

Acknowledgments

The authors would like to thank the 48 Centers of Excellence for their participation.

National Centers of Excellence: University of Arizona, Boston University Medical Center, Brown University/Women and Infants Hospital, University of California–Los Angeles, University of California–San Francisco, Drexel University, Harvard University/Brigham and Women's Hospital, University of Illinois–Chicago, Indiana University, Magee Women's Hospital University of Pittsburgh, University of Minnesota, University of Mississippi, University of Missouri, Oregon Health and Science University, University of Puerto Rico, University of Texas Health Sciences Center–San Antonio, Tulane and Xavier Universities of Louisiana, West Virginia University, University of Wisconsin–Madison, Virginia Commonwealth University, University of North Dakota, University of South Dakota, and University of Utah.

National Community Centers of Excellence: Christiana Care Health System, Great County Health Services, Griffin Hospital, Jefferson Health System, Kokua Kalihi Valley Comprehensive Family Services, Mariposa Community Health Center, Morton Plan Mease Health Care, North Point Health and Wellness Center, Northeastern Vermont Area Health Education Center, Northeast Missouri Health Council, North East Ohio Neighborhood Health Services, Oakhurst Medical Centers, St. Barnabas Hospital and Healthcare System, and Women's Health Services.

National Rural/Frontier Women's Health Coordinating Centers: Arizona Association of Community Health Centers, Colorado Coalition for the Homeless, Kentucky River Community Care, North Country Health Consortium, Rio Arriba Family Care Network, El Centro Family Health, Southern Area Health Education Center, Area Health Education Center of Southern Nevada, Utah Navajo Health System, Wisconsin Women's Health Foundation, Women's Wellness and Maternity Center, and Wyoming Health Council.

Disclosure Statement

No competing financial interests exist.

References

1. NIH Revitalization Act of 1993, Public Law 103–43. "NIH Guidelines on the Inclusion of Women and Minorities as Subjects in Clinical Research." March 18, 1994.
2. Bronfenbrenner U. The ecology of human development: experiments by nature and design. Cambridge, MA: Harvard University Press, 1979.
3. Multidisciplinary health models for women. The Office on Women's Health, 2008.
4. Dalkey N, Helmer O. An experimental application of the Delphi method to the use of experts. Management Science 1963;9:458–467.
5. Pinn VW. Interdisciplinary research on women's health and sex/gender factors. Gend Med 2005;2:121–123. Available at www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16290883.
6. Milliken N, Freund K, Pregler J, Reed S, Carlson K, Derman R, et al. Academic models of clinical care for women: the National Centers of Excellence in Women's Health. J Women's Health Gend Based Med 2001;10:627–636. Available at www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11571092.
7. Goodman RM, Seaver MR, Yoo S, Dibble S, Shada R, Sherman B, et al. A qualitative evaluation of the National Centers of Excellence in Women's Health Program. Women's Health Issues 2002;12:291–308. Available at www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12457571.
8. Centers for Disease Control. Costs of intimate partner violence against women in the United States. Atlanta, GA: US Department of Health and Human Services, CDC; 2003.
9. Campbell J, Jones AS, Dienemann J, Kub J, et al. Intimate partner violence and physical health consequences. Arch Intern Med 2002;162:1157–1163. Available at www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=12020187.
10. Jones AS, Dienemann J, Schollenberger J, et al. Long-term costs of intimate partner violence in a sample of female HMO enrollees. Women's Health Issues 2006;16:252–261. Available at www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=17055378.
11. Waters H, Hyder A, Rajkotia Y, Basu S, Rehwinkel J, Butchart A. The economic dimensions of interpersonal violence. Department of Injuries and Violence Prevention, World Health Organization, Geneva, 2004.

Address correspondence to:
Francisco A.R. Garcia, M.D., M.P.H.
1501 North Campbell Avenue
Tucson, AZ 85724

E-mail: fcisco@u.arizona.edu

Racial and Ethnic Differences in Hospice Use Among Patients With Heart Failure

Jane L. Givens, MD, MSCE; Jennifer Tjia, MD, MSCE; Chao Zhou, MA; Ezekiel Emanuel, MD, PhD; Arlene S. Ash, PhD

Background: Heart failure is the leading noncancer diagnosis for patients in hospice care and the leading cause of hospitalization among Medicare beneficiaries. Racial and ethnic differences in hospice patients are well documented for patients with cancer but poorly described for those with heart failure.

Methods: On the basis of a national sample of 98 258 Medicare beneficiaries 66 years and older on January 1, 2001, with a diagnosis of heart failure who had at least 1 physician or hospital encounter and who were not enrolled in hospice care between January 1 and December 31, 2000, we determined the effect of race and ethnicity on hospice entry for patients with heart failure in 2001 after adjusting for sociodemographic, clinical, and geographic factors.

Results: In unadjusted analysis, blacks (odds ratio [OR],0.52) and Hispanics (0.43) used hospice care for heart failure less than whites. Racial and ethnic differences in patients who received hospice care for heart failure persisted after adjusting for markers of income, urbanicity, severity of illness, local density of hospice use, and medical comorbidity (adjusted OR for blacks, 0.59; 95% confidence interval, 0.47-0.73; and adjusted OR for Hispanics, 0.49; 95% confidence interval, 0.37-0.66; compared with whites). Advanced age, greater comorbidity, emergency department visits, hospitalizations, and greater local density of hospice use were also associated with hospice use.

Conclusions: In a national sample of Medicare beneficiaries with heart failure, blacks and Hispanics used hospice care for heart failure less than whites after adjustment for individual and market factors. To understand the mechanisms underlying these findings, further examination of patient preferences and physician referral behavior is needed.

Arch Intern Med. 2010;170(5):427-432

Author Affiliations: Hebrew SeniorLife Institute for Aging Research (Dr Givens), Department of Medicine, Division of Gerontology, Beth Israel Deaconess Medical Center (Dr Givens), Health Care Research Unit, Boston University School of Medicine (Dr Ash and Ms Zhou), Boston, Massachusetts; Division of Geriatric Medicine, University of Massachusetts Medical School, Worcester (Dr Tjia); and Department of Bioethics, Clinical Center, National Institutes of Health, Bethesda, Maryland (Dr Emanuel).

HOSPICE CARE IS DESIGNED to provide comfort and emotional support to patients with terminal illness and their families and is most commonly (84%) provided through the Medicare hospice benefit.[1] Although patients dying of cancer comprised 44% of hospice admissions in 2006,[1] hospice care can also offer substantial benefit to patients with other terminal illnesses, such as end-stage heart failure (HF). However, patients in the terminal stages of noncancer conditions use hospice care less frequently than those with advanced malignant tumors.[1]

Currently, HF affects nearly 5 million people in the United States and is the leading cause of hospitalization among Medicare beneficiaries.[2] Despite progress in treatment,[3,4] patients with advanced HF have a 1-year mortality of 50% to 70%,[5] and hospice care is increasingly recommended in guidelines for such patients.[4,6-8] Although advanced heart disease represents the second most common hospice diagnosis, comprising 11.8% of hospice enrollees,[1] hospice services are generally recognized as underused by patients with HF.[4,9]

Underuse of hospice care is well documented, especially among racial and ethnic minorities.[10-14] Racial and ethnic differences in patients who use hospice care have been found across a spectrum of patients with cancer diagnoses[15-17] and may be more pronounced in patients with noncancer diagnoses.[18] However, previous studies[10,11,13-15,18] of ethnic differences in patients who use hospice care have been limited to cross-sectional or retrospective analyses, which have had limited ability to assess life expectancy or severity of illness among potential hospice care recipients. In this study, we use data from a large, nationally representative, ethnically diverse cohort of Medicare beneficiaries with HF to estimate the independent effect of race and ethnicity on hospice care for HF in the coming year, after adjusting for patient demographics, in-hospital interventions, comorbidity, and geographic variation in hospice use.

METHODS

DATA SOURCE

We used a national sample of Medicare beneficiaries oversampled for nonwhite patients as previously described.[19] The data source included merged Medicare claims files (denominator, inpatient, carrier, outpatient, and hospice files) from 2000 and 2001. An initial sample of 1 million beneficiaries 66 years and older was selected from the Medicare denominator file for 2001, with deliberate oversampling of beneficiaries who died in 2001 and racial and ethnic minorities categorized as black, Hispanic, and other. These files were merged with the National Death Index from 2001 to verify date of death.

STUDY POPULATION

We included Medicare beneficiaries who had complete claims data from 2000 and 2001, resided in the United States, were not enrolled in Medicare managed care organizations, and were not entitled to the Medicare end-stage renal disease benefit. Among those meeting these initial inclusion criteria (n=603 128), we limited the study sample to those beneficiaries (n=98 258) with a diagnosis of HF (*International Classification of Diseases, Ninth Revision, Clinical Modification* [*ICD-9-CM*][20] codes 428.xx, 398.91, 402.11, 402.91, 404.11, 404.13, 404.91, or 404.93) who had at least 1 physician or hospital encounter and who were not enrolled in hospice care between January 1 and December 31, 2000. On the basis of the sampling, these individuals reflect the experience of approximately 2.7 million beneficiaries.

MAIN OUTCOME MEASURES

Outcome Variable

Our primary outcome variable was entry into a hospice with an admitting diagnosis of HF between January 1 and December 31, 2001. We identified hospice entry from the first hospice admission date in 2001 and calculated hospice duration from first admission date until death or December 31, 2001, for nondecedents.

Predictor Variables

Race and Ethnicity. We used the Medicare denominator file categories of white, black, and Hispanic. We also collapsed all remaining categories (including Asian, North American natives, and unknown) into "other."

Medical Comorbidity. To characterize the morbidity of participants, we calculated prospective diagnostic cost group (DCG) scores derived from outpatient, inpatient, and carrier claims for the year 2000 (DxCG version 6.1 for SAS Windows; SAS Institute Inc, Cary, North Carolina). The DCG score predicts Medicare costs for the period designated as next year, as calculated from 1 year of *ICD-9-CM* diagnosis codes, age, and sex, and is expressed as a relative risk.[21] Thus, a score of 1.0 indicates an expected level of future health care use equal to the mean for Medicare beneficiaries, and 2.0 indicates expected costs that are twice as high. This score also predicts other outcomes, including mortality.[22]

Use of Medical Services and Other Sociodemographic Variables. To characterize each beneficiary's HF severity, we used the inpatient file from 2000 to calculate the number of emergency department (ED) visits and hospitalizations and the number of days spent in an intensive care unit (ICU) or coronary care unit (CCU). We used the denominator file to capture beneficiary age and sex and to define a geography-based socioeconomic status indicator (median income of zip code of residence) as determined from 2000 US Census data. We also used a marker for a state's Medicaid purchase (ie, Medicaid buy-in) of the Part B benefit as an indicator of low individual income.

Competing Hospice Diagnoses and Geographic Variables. We used our software's condition categories to identify other common morbidities present in 2000 that could lead to hospice care in 2001. These categories were cancer, dementia, stroke, and chronic obstructive pulmonary disease.

Because hospice availability[23] and use[4,10,24] differ by place of residence, we created 2 geographic variables from beneficiary zip code of residence. The first measures urbanicity because persons in rural settings typically have less access to hospice services.[10,23] We used the Beale rural-urban continuum codes[25] to categorize urbanicity as follows: metropolitan region with a population of 1 million or more, metropolitan region with a population of less than 1 million, nonmetropolitan region, and unknown.

We also developed a novel, hospice-specific, health service area (HSA)–based variable that we call *local hospice density*, derived from the larger data set of HF and non-HF beneficiaries who met our initial eligibility criteria (n=603 128). This variable describes the local prevalence of hospice use among the 2001 decedents in this data set (n=158 903). The HSAs are either single counties or clusters of counties that are relatively self-contained with respect to hospital care and delineate local health care markets for community-based primary inpatient care. We located each decedent in the HSA by his or her zip code of residence. We then defined the local density of hospice use as the percentage of the HSA's decedents who had entered into hospice care. For HSAs with fewer than 50 decedents in our data, we substituted the density of hospice use for the hospital referral region that contains it. Hospital referral regions consist of 1 or more HSAs and represent the tertiary market for medical care, including referral or specialty care.[26]

STATISTICAL ANALYSIS

We used descriptive statistics to characterize the demographics, health characteristics, health care use (including hospice entry), and mortality of the study population by racial or ethnic category. All analyses were conducted using sampling weights to obtain population-based estimates and a linearized variance estimator based on a first-order Taylor series linear approximation to compute standard errors.[27] The sample weights represent the reciprocal of the sampling probabilities for each of the 8 strata represented in the total sample of 1 million beneficiaries (2 based on decedent status and 4 based on race and ethnicity). To assess the statistical significance of bivariate associations, we used an adjusted Wald test for continuous variables, a design-based Pearson χ^2 test for categorical variables, and a nonparametric equality-of-medians test for medians.

To identify the independent association of race and ethnicity on hospice entry for HF, we developed a series of weighted logistic regression models with the outcome of hospice entry for HF in 2001. Models successively added covariate sets: race and ethnicity alone (model 1); age and sex (model 2); urbanicity, income, hospitalizations, ED visits, number of ICU and CCU days, diagnoses of cancer, chronic obstructive pulmonary disease, dementia, or stroke, and DCG score (model 3); and local hospice density (model 4). To examine whether our findings were attributable to differences in local health care delivery systems not captured in our data, we also developed a fixed-effects regression model that compares members of different racial and ethnic groups only when they reside in the

Table 1. Characteristics of the Study Participants in 2000 by Ethnicity[a]

Characteristic	All	White	Black	Hispanic	Other
Sample size	98 258	33 957	32 411	15 903	15 987
Population size[b]	2 724 200	2 404 426	231 337	38 987	49 450
Age, mean (linearized SE), y	79.9 (0.05)	80.1 (0.06)	78.2 (0.06)	77.7 (0.06)	79.6 (0.07)
Patients by age range (y) (weighted %)					
66-75	26.9	25.6	37.3	36.4	30.4
76-85	44.3	44.8	39.8	47.0	39.6
>85	28.9	29.6	22.9	16.7	30.0
Male sex, weighted %	39.0	39.6	32.8	39.7	40.8
Urbanicity, weighted %					
Metropolitan area ≥1 million	36.3	34.5	51.1	45.8	47.5
Metropolitan area <1 million	26.1	26.8	19.3	23.9	21.2
Nonmetropolitan	25.7	26.9	17.3	11.8	16.2
Unknown	12.0	11.7	12.4	18.5	15.1
Patients by median income by zip code of residence ($) (weighted %)					
<25 000	6.5	4.4	24.6	24.2	9.7
25 001-35 000	30.3	29.6	38.9	34.5	23.3
35 001-45 000	30.6	31.6	22.1	23.7	27.2
>45 000	32.5	34.3	14.3	17.6	39.8
Medicaid buy-in, weighted %	20.8	16.9	47.0	70.3	51.8
DCG score, mean (linearized SE)	2.5 (0.01)	2.5 (0.01)	2.6 (0.01)	2.8 (0.01)	2.6 (0.01)
Any hospitalization, weighted %	54.7	54.6	56.1	57.3	49.4
No. of hospitalizations, mean (linearized SE)	1.06 (0.01)	1.05 (0.01)	1.16 (0.01)	1.23 (0.01)	0.94 (0.01)
Any CCU stay, weighted %	11.8	11.9	11.5	14.1	10.7
Days in CCU, mean (linearized SE)	0.65 (0.02)	0.65 (0.02)	0.64 (0.02)	0.88 (0.03)	0.57 (0.02)
Any ICU stay, weighted %	19.3	19.3	18.6	24.3	19.1
Days in ICU, mean (linearized SE)	1.04 (0.03)	1.02 (0.03)	1.11 (0.03)	1.59 (0.04)	1.13 (0.03)
Any ED visit, weighted %	48.2	48.3	46.6	54.2	45.5
No. of ED visits, mean (linearized SE)	1.06 (0.01)	1.05 (0.01)	1.12 (0.01)	1.39 (0.02)	0.97 (0.01)

Abbreviations: CCU, coronary care unit; DCG, diagnostic cost group; ED, emergency department; ICU, intensive care unit; NA, not available.
[a]Statistical significance by adjusted Wald test for continuous variables and Pearson χ^2 statistic for categorical variables ($P<.001$ for all entries).
[b]Estimated from sampling weights.

same HSA. Because the findings of this sensitivity analysis are similar to the more conceptually revealing original analysis described herein, we present the results of the original analysis.

Analyses were conducted using STATA statistical software, version 10.0 (StataCorp, College Station, Texas), and SAS statistical software, version 9.0 (SAS Institute Inc). This study was approved by the institutional review board of the Boston University School of Medicine.

RESULTS

CHARACTERISTICS OF THE STUDY POPULATION

The final study population included 98 258 Medicare beneficiaries (weighted population, n=2.7 million). The weighted percentages by race and ethnicity were as follows: white, 88.3%; black, 8.5%; Hispanic, 1.4%; and other, 1.8%. Many differences in sociodemographic characteristics, DCG scores, and severity of illness measures by race and ethnicity were statistically significant (**Table 1**). Beneficiaries who were of black, Hispanic, and other races and ethnicities were more likely to live in large metropolitan urban areas. Black and Hispanic beneficiaries had lower incomes and were more likely to have Medicaid buy-in. There were also notable age differences among the groups, with black and Hispanic beneficiaries being younger than whites and others. Differences in markers of health and health care use were notable, with Hispanics having the highest mean DCG score, number of hospitalizations, length of stays in the CCU and ICU, and number of ED visits. Unadjusted use of hospice care for HF and for any reason differed by race or ethnicity (**Table 2**). The percentage of beneficiaries using hospice care for any diagnosis in 2001 was small (3.9%). Of these, 18.2% entered because of HF. The percentage of decedents who had used hospice care was 19.9% overall, with a higher percentage of white decedents (20.4%) using hospice care compared with blacks (15.4%), Hispanics (16.9%), or those of other races or ethnicities (16.3%). Among those who used hospice care, a higher percentage of whites entered hospice care for HF treatment (18.5%) than blacks (14.1%), Hispanics (13.2%), and other racial and ethnic groups (15.8%). For HF, Hispanics had the longest median duration of hospice care (19 days), followed by blacks (14 days), whites (13 days), and other racial and ethnic groups (10 days).

MULTIVARIABLE MODELS FOR HOSPICE USE FOR HF

Nonwhite patients used hospice care for HF less than whites (**Table 3**). After adjusting for sociodemographics, urbanicity, comorbidities, DCG score, use of medical services, and local hospice density (model 4), hospice use remained lowest for Hispanic beneficiaries (adjusted odds ratio [aOR], 0.49) compared with whites.

Table 2. Mortality and Hospice Use in 2001 by Patient Ethnicity[a]

	All	White	Black	Hispanic	Other
Sample size	98 258	33 957	32 411	15 903	15 987
Population size[b]	2 724 200	2 404 426	231 337	38 987	49 450
Death, weighted %	15.9	16.1	14.7	11.7	14.1
Any hospice use, weighted %	3.9	4.1	2.8	2.4	2.8
Decedents who used hospice care, weighted %	19.9	20.4	15.4	16.9	16.3
Hospice users with an admitting diagnosis of heart failure, weighted %	18.2	18.5	14.1	13.2	15.8
Days of hospice use (all diagnoses), median (25th-75th percentile)	12 (5-37)	11 (5-37)	14 (5-40)	12 (4-32)	11 (5-33)
Days of hospice use (heart failure), median (25th-75th percentile)	13 (4-38)	13 (5-38)	14 (4-41)	19 (6-58)	10 (4-34)
Local hospice use density, mean (SE)[c]	25.5 (0.001)	25.8 (0.001)	24.0 (0.001)	23.3 (0.001)	23.1 (0.001)

[a] Statistical significance by Pearson χ^2 statistic for categorical variables and nonparametric equality-of-medians test ($P < .001$ for all entries).
[b] Estimated from sampling weights.
[c] Mean percentage of all decedents who used hospice care in each beneficiary's health service area.[26]

Table 3. Year 2000 Predictors of Entry Into a Hospice With an Admitting Diagnosis of Heart Failure in 2001

	Odds Ratios (95% Confidence Intervals)			
	Model 1	Model 2[a]	Model 3[a]	Model 4[a]
Ethnicity				
White	1 [Reference]	1 [Reference]	1 [Reference]	1 [Reference]
Black	0.52 (0.44-0.63)	0.60 (0.50-0.72)	0.57 (0.46-0.71)	0.59 (0.47-0.73)
Hispanic	0.43 (0.34-0.54)	0.53 (0.41-0.67)	0.46 (0.34-0.61)	0.49 (0.37-0.66)
Other	0.58 (0.48-0.70)	0.59 (0.48-0.71)	0.57 (0.46-0.71)	0.64 (0.52-0.80)
Age, y		1.08 (1.07-1.09)	1.09 (1.07-1.10)	1.09 (1.07-1.10)
Male sex		1.17 (0.95-1.43)	1.17 (0.95-1.45)	1.16 (0.94-1.44)
Urbanicity				
Metropolitan area ≥1 million			1 [Reference]	1 [Reference]
Metropolitan area <1 million			1.07 (0.83-1.38)	1.08 (0.84-1.40)
Nonmetropolitan			0.82 (0.62-1.09)	0.90 (0.68-1.20)
Unknown			1.19 (0.87-1.63)	1.07 (0.78-1.48)
Median income by zip code of residence, $				
<25 000			1 [Reference]	1 [Reference]
25 001-35 000			1.00 (0.64-1.57)	0.92 (0.58-1.44)
35 001-45 000			0.97 (0.61-1.54)	0.84 (0.53-1.34)
>45 000			1.16 (0.73-1.86)	0.98 (0.61-1.57)
Medicaid buy-in in 2000 (vs no buy-in)			0.99 (0.76-1.30)	1.03 (0.79-1.34)
Diagnostic cost group score			1.11 (1.03-1.20)	1.13 (1.04-1.22)
Health care use				
Emergency department visits			1.04 (0.99-1.09)	1.05 (1.00-1.10)
No. of hospitalizations			1.14 (1.07-1.21)	1.13 (1.07-1.20)
Intensive care unit, d			1.00 (0.98-1.01)	1.00 (0.98-1.01)
Coronary care unit, d			1.02 (1.00-1.03)	1.01 (1.00-1.03)
Competing hospice diagnoses				
Cancer (any type)			0.70 (0.57-0.87)	0.72 (0.58-0.89)
COPD			1.17 (0.95-1.45)	1.14 (0.92-1.41)
Dementia			1.07 (0.82-1.39)	1.07 (0.83-1.39)
Stroke			0.98 (0.75-1.29)	0.98 (0.74-1.29)
Local density of hospice use				
Lowest quintile				1 [Reference]
Second quintile				2.54 (1.69-3.80)
Third quintile				3.27 (2.28-4.69)
Fourth quintile				4.30 (3.05-6.07)
Highest quintile				6.10 (4.37-8.53)

Abbreviation: COPD, chronic obstructive pulmonary disease.
[a] Odds ratios were adjusted for models 2, 3, and 4.

Blacks (aOR, 0.59) and other nonwhite beneficiaries (aOR, 0.64) with HF were also less likely to use hospice care. Sex, income, Medicaid buy-in status, more days in the ICU or CCU, greater number of ED visits, and geographic residence in an urban area were not significantly associated with hospice entry. Higher DCG score, advanced age, ED visits and hospitalizations, and greater local hospice density were associated with hospice use. Competing hospice diagnoses of chronic obstructive pulmonary disease, dementia, and stroke were not associated with hospice use, but a diagnosis of cancer was protective against entry into hospice care for HF (aOR, 0.72).

COMMENT

This analysis of a national cohort of Medicare beneficiaries finds marked racial and ethnic differences among patients with HF who received hospice care. In our sample, blacks had 41.0% lower odds of hospice use and Hispanics 51.0% lower odds of receiving hospice care than whites. To our knowledge, our study is the largest longitudinal cohort study examining factors associated with hospice entry for a noncancer diagnosis. In addition, we use a prospective design and a novel measure of local hospice use.

Our finding that racial and ethnic minorities are less likely to use hospice care for HF is consistent with other hospice literature. Several studies have documented lower rates of hospice use among blacks, Hispanics, and Asian Americans compared with whites. However, these studies either have specifically focused on patients with cancer[14-16] or do not define the diagnosis determining hospice entry.[12,13,28] There is concern that racial and ethnic disparities may be more pronounced among patients with noncancer diagnoses than among those with cancer.[18] Our findings document significant racial and ethnic differences in hospice use and counter speculation that overall increases in the availability of hospice services in the 1990s may have erased racial and ethnic differences in hospice use.[11]

In this study, we have adjusted for numerous factors related to hospice entry, many of which vary substantially by race and ethnicity. Geographic variation in hospice availability is associated with patterns of hospice use,[10] including hospice entry for HF.[4] Hospice availability differs between urban and rural areas,[10,23,24] and lower hospice use has been documented among patients living in predominantly minority vs white areas.[29] On a state level, greater regional availability of hospice services is associated with less racial and ethnic disparity between blacks and whites in hospice use.[28] We created a variable to represent the prevalence of hospice use among Medicare decedents within smaller local geographic units (HSAs) and found that our observed ethnic differences in hospice use for HF persisted after accounting for local hospice availability and urbanicity.

In addition to sociodemographic, clinical, and geographic characteristics, cultural beliefs and values may contribute to differences between blacks and whites in end-of-life care and hospice use.[30,31] For example, compared with whites, blacks are less likely to complete advance directives, have less favorable beliefs about hospice care, opt for more aggressive treatments, and are more likely to have spiritual beliefs that conflict with the goals of palliative treatment.[19,30,32] In addition, lack of trust between patients and physicians may be more pronounced for ethnic minorities[33,34] and may contribute to ethnic differences in hospice entry.[35,36] Blacks more often report receiving inadequate information regarding end-of-life care[37] and are less likely to be informed about hospice services than whites.[38] Our administrative data contained information on neither patient cultural beliefs and values nor physician behavior, factors that may also help explain differences in hospice use. Finally, health literacy, also not measured in this study, has been found to partially explain racial and ethnic differences in end-of-life treatment preferences.[39]

This study has several limitations. First, we could not measure HF severity using clinical markers such as ejection fraction or exercise tolerance, and thus, HF severity might vary by race and ethnicity. However, we adjusted for prior-year measures of ED visits, hospitalizations, and ICU and CCU use as proxies for clinical complications and disease severity. Second, although race and ethnicity coding in the Medicare denominator file has a good positive predictive value for identifying whites, blacks, and Hispanics (>94%), it performs less well for Asians and American Indian/Alaskan Native beneficiaries,[40] making findings for the "other" race and ethnicity category difficult to interpret. Finally, although higher hospice use has been reported among enrollees from health maintenance organizations compared with patients with traditional Medicare coverage,[41] we could not include such patients in our study sample because of incomplete capture of diagnoses and health care use.

This article describes the largest noncancer study of hospice entry to date. It prospectively examines a national probability sample of people with HF, oversampled for racial and ethnic minorities and decedents. Large differences in racial and ethnic minority use of hospice care for HF compared with whites remain largely unchanged after adjusting for differences in income, urbanicity, comorbidity, severity of illness, and hospice use density. It is not clear how many of these differences reflect access issues as opposed to considered patient preferences.

Accepted for Publication: September 25, 2009.
Correspondence: Jane L. Givens, MD, MSCE, Hebrew SeniorLife Institute for Aging Research, Beth Israel Deaconess Medical Center, 1200 Centre St, Boston, MA 02131 (janegivens@hrca.harvard.edu).
Author Contributions: *Study concept and design*: Givens, Tjia, and Ash. *Acquisition of data*: Emanuel and Ash. *Analysis and interpretation of data*: Givens, Tjia, Zhou, and Ash. *Drafting of manuscript*: Givens and Tjia. *Critical revision of manuscript for important intellectual content*: Givens, Tjia, Zhou, Emanuel, and Ash. *Statistical analysis*: Givens, Tjia, Zhou, and Ash. *Obtained funding*: Emanuel. *Study supervision:* Givens and Ash.
Financial Disclosure: Dr Ash is cofounder of and consulting senior scientist at DxCG Inc (Boston, Massachusetts), a company whose software was used to measure morbidity burden in this study. Otherwise, there are no conflicts of interest or financial disclosures.
Funding/Support: This study was supported by contracts with the Department of Bioethics of the Clinical Center of the National Institutes of Health. Dr Tjia is supported by National Institute on Aging Career Development Award K08AG021527.
Role of the Sponsors: The funding sources had no role in the design and conduct of the study; in the collection, management, analysis, and interpretation of the data; or in the preparation of the manuscript.
Additional Information: Drs Emanuel and Ash arranged for use of Medicare and Medicaid Services data.

REFERENCES

1. National Hospice and Palliative Care Organization. *NHCPO Facts and Figures: Hospice Care in America.* Alexandria, VA: National Hospice and Palliative Care Organization; 2008.
2. Thom T, Haase N, Rosamond W, et al; American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Heart disease and stroke statistics—2006 update. *Circulation.* 2006;113(6):e85-e151.
3. Kosiborod M, Lichtman JH, Heidenreich PA, et al. National trends in outcomes among elderly patients with heart failure. *Am J Med.* 2006;119(7):616.e1-616.e7.
4. Hauptman PJ, Goodlin SJ, Lopatin M, Costanzo MR, Fonarow GC, Yancy CW. Characteristics of patients hospitalized with acute decompensated heart failure who are referred for hospice care. *Arch Intern Med.* 2007;167(18):1990-1997.
5. Wolinsky FD, Smith DM, Stump TE, Overhage JM, Lubitz RM. The sequelae of hospitalization for congestive heart failure among older adults. *J Am Geriatr Soc.* 1997;45(5):558-563.
6. Goodlin SJ, Hauptman PJ, Arnold R, et al. Consensus statement: palliative and supportive care in advanced heart failure. *J Card Fail.* 2004;10(3):200-209.
7. Hunt SA, Baker DW, Chin MH, et al; American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee to Revise the 1995 Guidelines for the Evaluation and Management of Heart Failure); International Society for Heart and Lung Transplantation; Heart Failure Society of America. ACC/AHA Guidelines for the Evaluation and Management of Chronic Heart Failure in the Adult: executive summary: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee to Revise the 1995 Guidelines for the Evaluation and Management of Heart Failure): developed in collaboration with the International Society for Heart and Lung Transplantation; endorsed by the Heart Failure Society of America. *Circulation.* 2001;104(24):2996-3007.
8. Hauptman PJ, Havranek EP. Integrating palliative care into heart failure care. *Arch Intern Med.* 2005;165(4):374-378.
9. Goodlin SJ, Kutner JS, Connor SR, Ryndes T, Houser J, Hauptman PJ. Hospice care for heart failure patients. *J Pain Symptom Manage.* 2005;29(5):525-528.
10. Virnig BA, Kind S, McBean M, Fisher E. Geographic variation in hospice use prior to death. *J Am Geriatr Soc.* 2000;48(9):1117-1125.
11. Han B, Remsburg RE, Iwashyna TJ. Differences in hospice use between black and white patients during the period 1992 through 2000. *Med Care.* 2006;44 (8):731-737.
12. Enguidanos S, Yip J, Wilber K. Ethnic variation in site of death of older adults dually eligible for Medicaid and Medicare. *J Am Geriatr Soc.* 2005;53(8):1411-1416.
13. Greiner KA, Perera S, Ahluwalia JS. Hospice usage by minorities in the last year of life: results from the National Mortality Followback Survey. *J Am Geriatr Soc.* 2003;51(7):970-978.
14. Colón M, Lyke J. Comparison of hospice use and demographics among European Americans, African Americans, and Latinos. *Am J Hosp Palliat Care.* 2003; 20(3):182-190.
15. Virnig BA, Marshall McBean A, Kind S, Dholakia R. Hospice use before death: variability across cancer diagnoses. *Med Care.* 2002;40(1):73-78.
16. Ngo-Metzger Q, Phillips RS, McCarthy EP. Ethnic disparities in hospice use among Asian-American and Pacific Islander patients dying with cancer. *J Am Geriatr Soc.* 2008;56(1):139-144.
17. Smith AK, Earle CC, McCarthy EP. Racial and ethnic differences in end-of-life care in fee-for-service Medicare beneficiaries with advanced cancer. *J Am Geriatr Soc.* 2009;57(1):153-158.
18. Johnson KS, Kuchibhatla M, Tanis D, Tulsky JA. Racial differences in the growth of noncancer diagnoses among hospice enrollees. *J Pain Symptom Manage.* 2007; 34(3):286-293.
19. Hanchate A, Kronman AC, Young-Xu Y, Ash AS, Emanuel E. Racial and ethnic differences in end-of-life costs: why do minorities cost more than whites? *Arch Intern Med.* 2009;169(5):493-501.
20. World Health Organization. *International Classification of Diseases, Ninth Revision, Clinical Modification.* Geneva, Switzerland: World Health Organization; 2009.
21. Ash AS, Ellis RP, Pope GC, et al. Using diagnoses to describe populations and predict costs. *Health Care Financ Rev.* 2000;21(3):7-28.
22. Ash AS, Posner MA, Speckman J, Franco S, Yacht AC, Bramwell L. Using claims data to examine mortality trends following hospitalization for heart attack in Medicare. *Health Serv Res.* 2003;38(5):1253-1262.
23. Virnig BA, Ma H, Hartman LK, Moscovice I, Carlin B. Access to home-based hospice care for rural populations: identification of areas lacking service. *J Palliat Med.* 2006;9(6):1292-1299.
24. Virnig BA, Moscovice IS, Durham SB, Casey MM. Do rural elders have limited access to Medicare hospice services? *J Am Geriatr Soc.* 2004;52(5):731-735.
25. US Department of Agriculture. *Measuring Rurality: Rural-Urban Continuum Codes.* Washington, DC: US Dept of Agriculture; 2006.
26. Wennberg J, Cooper M, Birkmeyer JD, et al. *The Dartmouth Atlas of Health Care.* Hanover, NH: Dartmouth Medical School, Center for Evaluative Clinical Sciences; 1998.
27. Korn E, Graubard B. *Analysis of Health Surveys.* New York, NY: John Wiley & Sons Inc; 1999.
28. Connor SR, Elwert F, Spence C, Christakis NA. Racial disparity in hospice use in the United States in 2002. *Palliat Med.* 2008;22(3):205-213.
29. Haas JS, Earle CC, Orav JE, et al. Lower use of hospice by cancer patients who live in minority versus white areas. *J Gen Intern Med.* 2007;22(3):396-399.
30. Johnson KS, Kuchibhatla M, Tulsky JA. What explains racial differences in the use of advance directives and attitudes toward hospice care? *J Am Geriatr Soc.* 2008;56(10):1953-1958.
31. Johnson KS, Kuchibhatla M, Tanis D, Tulsky JA. Racial differences in hospice revocation to pursue aggressive care. *Arch Intern Med.* 2008;168(2):218-224.
32. Steinhauser KE, Christakis NA, Clipp EC, McNeilly M, McIntyre L, Tulsky JA. Factors considered important at the end of life by patients, family, physicians, and other care providers. *JAMA.* 2000;284(19):2476-2482.
33. Morrison RS, Zayas LH, Mulvihill M, Baskin SA, Meier DE. Barriers to completion of health care proxies: an examination of ethnic differences. *Arch Intern Med.* 1998;158(22):2493-2497.
34. Blackhall LJ, Frank G, Murphy ST, Michel V, Palmer JM, Azen SP. Ethnicity and attitudes towards life sustaining technology. *Soc Sci Med.* 1999;48(12):1779-1789.
35. Krakauer EL, Crenner C, Fox K. Barriers to optimum end-of-life care for minority patients. *J Am Geriatr Soc.* 2002;50(1):182-190.
36. Reese DJ, Ahern RE, Nair S, O'Faire JD, Warren C. Hospice access and use by African Americans: addressing cultural and institutional barriers through participatory action research. *Soc Work.* 1999;44(6):549-559.
37. Welch LC, Teno JM, Mor V. End-of-life care in black and white: race matters for medical care of dying patients and their families. *J Am Geriatr Soc.* 2005;53 (7):1145-1153.
38. Rhodes RL, Teno JM, Welch LC. Access to hospice for African Americans: are they informed about the option of hospice? *J Palliat Med.* 2006;9(2):268-272.
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. McBean A. *Improving Medicare's Data on Race and Ethnicity.* Washington, DC: National Academy of Social Insurance; 2006.
41. Virnig BA, Persily NA, Morgan RO, DeVito CA. Do Medicare HMOs and Medicare FFS differ in their use of the Medicare hospice benefit? *Hosp J.* 1999;14(1):1-12.

Effectiveness of Policies Restricting Hours of Alcohol Sales in Preventing Excessive Alcohol Consumption and Related Harms

Robert A. Hahn, PhD, MPH, Jennifer L. Kuzara, MA, MPH, Randy Elder, PhD, Robert Brewer, MD, MSPH, Sajal Chattopadhyay, PhD, Jonathan Fielding, MD, MPH, MBA, Timothy S. Naimi, MD, MPH, Traci Toomey, PhD, Jennifer Cook Middleton, PhD, Briana Lawrence, MPH, the Task Force on Community Preventive Services

Abstract: Local, state, and national policies that limit the hours that alcoholic beverages may be available for sale might be a means of reducing excessive alcohol consumption and related harms. The methods of the *Guide to Community Preventive Services* were used to synthesize scientific evidence on the effectiveness of such policies. All of the studies included in this review assessed the effects of increasing hours of sale in on-premises settings (in which alcoholic beverages are consumed where purchased) in high-income nations. None of the studies was conducted in the U.S. The review team's initial assessment of this evidence suggested that changes of less than 2 hours were unlikely to significantly affect excessive alcohol consumption and related harms; to explore this hypothesis, studies assessing the effects of changing hours of sale by less than 2 hours and by 2 or more hours were assessed separately.

There was sufficient evidence in ten qualifying studies to conclude that increasing hours of sale by 2 or more hours increases alcohol-related harms. Thus, disallowing extensions of hours of alcohol sales by 2 or more should be expected to prevent alcohol-related harms, while policies decreasing hours of sale by 2 hours or more at on-premises alcohol outlets may be an effective strategy for preventing alcohol-related harms. The evidence from six qualifying studies was insufficient to determine whether increasing hours of sale by less than 2 hours increases excessive alcohol consumption and related harms.

(Am J Prev Med 2010;39(6):590–604) Published by Elsevier Inc. on behalf of American Journal of Preventive Medicine

Introduction

Excessive alcohol consumption is responsible for approximately 79,000 deaths per year in the U.S., making it the third-leading cause of preventable death.[1] Binge drinking (consuming five or more drinks per occasion for men and four or more drinks per occasion for women) is reported by approximately 15% of U.S. adults aged ≥18 years and by approximately 29% of high school students in the U.S.[2,3] The direct and indirect economic costs of excessive drinking in 1998 were $184.6 billion.[4] The reduction of excessive alcohol consumption in general and binge drinking in particular are thus matters of major public health and economic interest. Reducing binge drinking among U.S. adults has been a public health objective in *Healthy People 2010*.[5]

In the U.S., local control of the total or specific hours during which alcoholic beverages may be sold (hereafter referred to as "hours of sale") varies from one state to another. Some states allow cities, counties, and other local jurisdictions to enact their own alcohol control policies, and in these states, restrictions on hours of sale can vary from one location to another. In other states, local control may be pre-empted by state regulations that prohibit local authorities from enacting alcohol control regulations stricter than those that apply to the rest of the state.[6,7] As of 1953, American Indian reservations have the authority to establish their own alcohol-related policies, prior to which alcohol was formally prohibited.[8]

There is also wide variation among states in the specific restrictions they place on the hours of sale by retail setting (i.e., on- or off-premises) and by the day of the week.[9] For on-premises alcohol outlets, states allow facilities to serve alcohol for a median of 19 hours a day on weekdays and

From the Community Guide Branch of the Epidemiology and Analysis Program Office (Hahn, Kuzara, Elder, Chattopadhyay, Middleton, Lawrence), National Center for Chronic Disease Prevention and Health Promotion (Brewer, Naimi), CDC, Atlanta, Georgia; Los Angeles County Department of Public Health (Fielding), Los Angeles, California; University of Minnesota School of Public Health (Toomey), Minneapolis, Minnesota

The names and affiliations of the Task Force members are listed at www.thecommunityguide.org/about/task-force-members.html.

Address correspondence to: Robert A. Hahn, PhD, MPH, Community Guide Branch, Epidemiology and Analysis Program Office, CDC, 1600 Clifton Road, Mailstop E-69, Atlanta GA 30333. E-mail: rhahn@cdc.gov.

0749-3797/$17.00

doi: 10.1016/j.amepre.2010.09.016

Saturdays. Nine states (Alabama, Florida, Georgia, Illinois, Louisiana, Maryland, Nevada, New Jersey, and South Carolina) have no limits on hours of sale for on-premises alcohol outlets.[9] On Sundays, alcohol may be served for a median of 17 hours at on-premises facilities, with seven states placing no restrictions on Sunday on-premises sales; four states allow no sales of alcohol at on-premises facilities on Sundays. In off-premises settings, hours of sale are limited to a median of 18 hours on weekdays and Saturdays. Restrictions range from no limits on hours of sale in Alabama, Florida, Georgia, Illinois, Louisiana, Maryland, and Nevada to 8 hours of sale allowed in Idaho. On Sundays, states allow a median of 13 hours of alcohol sales at off-premises facilities, with five states having no restrictions; 18 states with "blue laws" allow no off-premises sales.

This review uses the methods of the *Guide to Community Preventive Services (Community Guide)*[10] to assess the effects of changes in the hours during which alcohol is served on excessive alcohol consumption and related harms. A separate review published in this issue assesses the effects of changing days of sale on excessive alcohol consumption and related harms and concludes that increasing days of sale leads to increased consumption and related harms. The focal question of the present review is how, within allowable days of sale, the number of hours available for acquisition and service of alcohol affects excessive alcohol consumption and related harms.

Findings and Recommendations from Other Reviews and Advisory Groups

Several scientific reviews[11–14] have concluded that restricting the hours when alcohol may be sold is an effective strategy for reducing excessive alcohol consumption and related harms. One review,[11] funded by the Center for Substance Abuse Prevention (CSAP), found substantial evidence of harms associated with expanding the hours and days of alcohol sales. This conclusion was based on previous empirical research indicating that the expansion of the hours and days of sale increased prevalence of excessive alcohol consumption and alcohol-related problems. Most prior reviews have combined findings on days and hours and none have examined a threshold effect. The CSAP review included studies prior to 1999; a recent review[14] includes studies published between 2000 and 2008. The present review covers both periods using the systematic methods of the *Community Guide* described below.

Several international bodies have also recommended the control of hours or days of sale, or both as means of reducing excessive alcohol consumption and related harms.[15] For example, a recent review[16] of alcohol control strategies by the WHO found that limiting of hours of sale was an effective method for reducing alcohol-related harms. In Ireland, the Department of Health and Children's Strategic Task Force on Alcohol[17] concluded (p. 30) that "restricting any further increases in the physical availability of alcohol (number of outlets and times of sales)" is among the most effective policy measures for influencing alcohol consumption and related harms.

Methods

The methods of the *Community Guide* were used to systematically review scientific studies that have evaluated the effectiveness of limiting or maintaining existing limits on the hours of sale for preventing excessive alcohol consumption and related harms.[10] In brief, the *Community Guide* process involves forming a systematic review development team (review team), consisting of subject matter and methodology experts from other parts of the CDC, other federal agencies, and academia, and the Task Force on Community Preventive Services (Task Force); developing a conceptual approach for organizing, grouping, and selecting interventions; selecting interventions to evaluate; searching for and retrieving available research evidence on the effects of those interventions; assessing the quality of and abstracting information from each study that meets inclusion criteria; assessing the quality of and drawing conclusions about the body of evidence on intervention effectiveness; and translating the evidence on effectiveness into recommendations. Evidence is collected and summarized on (1) the effectiveness of reviewed interventions in altering selected health-related outcomes and (2) positive or negative effects 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 of evidence (i.e., the extent to which available effectiveness data might generalize to diverse population segments and settings); (4) barriers to implementation; and (5) the economic impact of the intervention. To help ensure objectivity, the review process is typically led by scientists who are not employed by a program that might be responsible for overseeing the implementation of the intervention being evaluated.

The results of this review process are then presented to the Task Force, an independent scientific review board that objectively considers the scientific evidence on intervention effectiveness presented to them and then determines, with the guidance of a translation table, whether the evidence is sufficient to warrant a recommendation on intervention effectiveness.[10] Evidence can be found to be strong, sufficient, or insufficient. Sufficient or strong evidence may indicate benefit, harm, or ineffectiveness of the intervention whereas insufficient evidence indicates more research is needed.

Conceptual Approach and Analytic Framework

The premise of this review is that increased availability of alcoholic beverages through any mechanism facilitates increases in excessive consumption and related harms, and that limiting hours of sale of alcoholic beverages is one way to reduce availability. The limitation of hours of sale of alcoholic beverages was defined as "applying regulatory authority to limit the hours that alcoholic beverages may be sold at on- and off-premises alcoholic beverage outlets." *Limiting* may refer to either *maintaining existing limits* in response to efforts to expand hours of sale or *reducing current limits on hours of sale.* Hours of sale may be regulated at the national, state, or local level or some combination of these. *Off-premises retailing* refers to

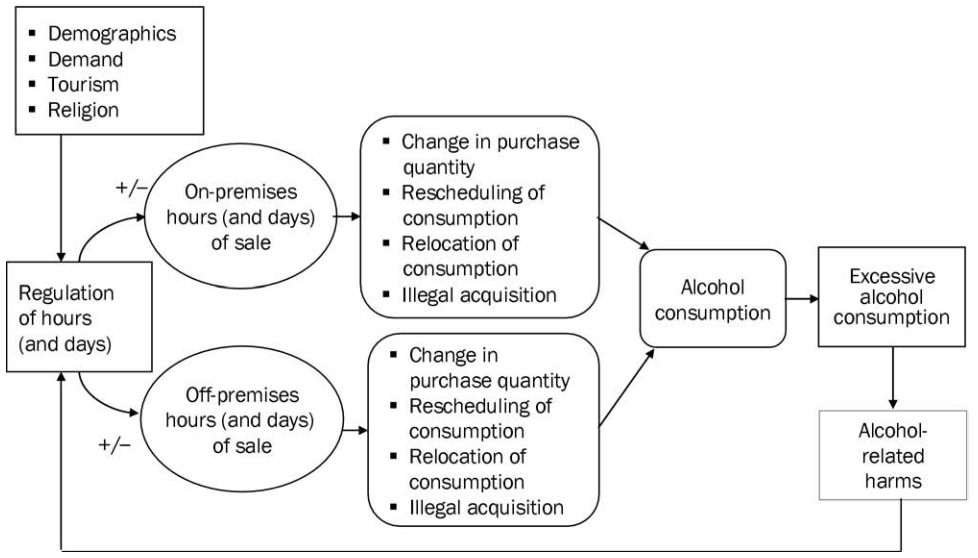


Figure 1. Effects of regulation of hours (and days) of alcohol sales on excessive alcohol consumption and related harms

the sale of contained alcoholic beverages, for instance, at package stores, liquor stores, grocery stores, or convenience stores, for consumption elsewhere. *On-premises retailing* refers to the sale of alcoholic beverages for consumption at the point of sale, for example, at bars, restaurants, or clubs.

Policies that regulate the hours of sale may be influenced by various characteristics of the affected population, including the demand for alcoholic beverages, the age distribution of the population, the religious affiliation and involvement of residents, and the amount of tourism the area attracts. Policies reducing or expanding hours of sale are hypothesized to affect alcohol consumption and alcohol-related harms through the following means (Figure 1). First, increases or decreases in the hours of sale affect consumers' ability to purchase alcohol by changing its availability. Second, when access to alcoholic beverages changes, consumers may alter their purchasing habits in several ways, including changing their purchase volume, rescheduling their purchases, relocating their purchases, or obtaining alcoholic beverages illegally. Changes in their purchasing habits may then affect their drinking patterns or overall levels of alcohol use, resulting in changes in alcohol-related problems.

Changes in the hours of sale may also affect alcohol-related health outcomes by other means. For example, increases in the hours that alcohol is available at on-premises outlets may be associated with increased social aggregation, which, in turn, may increase aggressive behaviors that are exacerbated by alcohol consumption.[18] Increases or decreases in the hours that alcohol is available in one jurisdiction may also increase or decrease alcohol consumption in adjacent jurisdictions if consumers travel from a jurisdiction with fewer hours to one with greater hours. This may also affect the number of miles traveled to purchase alcohol, and therefore the probability of alcohol-related motor vehicle crashes.

The present review addresses the following research question: what are the effects on excessive alcohol consumption and related harms of changing the hours of sale at on- or off-premises outlets? It was hypothesized that there would be a dose–response relationship related to the magnitude of the change in hours (i.e., the amount by which hours of sale are increased or decreased). Based on this hypothesis, the body of evidence for this review was stratified into studies examining changes of ≥2 hours and <2 hours per day. This cut point was chosen by the judgment of the review team that 2 hours might be a reasonable threshold for a substantial effect and on the distribution of available studies.

The process by which hours of alcohol sale are changed in different settings may also be an important variable to consider in evaluating the effects of such changes. In some settings in which the allowable hours of sale are increased, any licensed facility may extend hours. In others, facilities must apply for an extension and meet certain criteria, such as demonstrating a lack of facility crowding in a neighborhood. It was hypothesized that the additional level of regulation required to apply for extended opening hours might reduce the potential harm from greater access by restricting the implementation and extent of added hours.

Inclusion and Exclusion Criteria

To be included as evidence in this review, studies had to meet certain criteria. First, studies that assessed short-term changes in alcohol availability (e.g., alcohol sales related to a special event such as a sports competition) were not included. Second, eligible studies needed to assess the specific impact of changes in the hours of sale on excessive alcohol consumption, related harms, or both, as opposed to evaluating the effect of change in combination with other interventions. Studies of combined interventions may obscure the effects attributable specifically to changes in hours. Third, because the current focus was on the effects of changes in hours of sale in jurisdictions where these changes occurred, no review was made of studies that examined the effects of changes in hours in one jurisdiction on consumption elsewhere, for example, in neighboring jurisdictions or across a border. Fourth, to increase the applicability of the findings to the U.S., studies had to be conducted in countries with high-income economies[a] according to the World Bank.[19] Fifth, studies had to present primary research findings, not just review other research findings. Sixth, studies had to be published in English. Seventh, studies had to have a comparison group

[a]World Bank High-Income Economies (as of May 5, 2009): Andorra, Antigua and Barbuda, Aruba, Australia, Austria, The Bahamas, Bahrain, Barbados, Belgium, Bermuda, Brunei Darussalam, Canada, Cayman Islands, Channel Islands, Cyprus, Czech Republic, Denmark, Equatorial Guinea, Estonia, Faeroe Islands, Finland, France, French Polynesia, Germany, Greece, Greenland, Guam, Hong Kong (China), Hungary, 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, Northern Mariana Islands, Norway, Oman, Portugal, Puerto Rico, Qatar, San Marino, Saudi Arabia, Singapore, Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Trinidad and Tobago, United Arab Emirates, United Kingdom, U.S., Virgin Islands (U.S.).

or, at a minimum, compare outcomes of interest before and after a change in the policy related to hours of sale.

Specific types of alcohol-related harms of interest were alcohol-related diseases (e.g., liver cirrhosis), alcohol-impaired driving, alcohol-related crashes, unintentional or intentional injuries, and violent crime. When studies assessed multiple outcomes of interest, those outcomes with the strongest known association with excessive alcohol consumption were selected. Outcome measures that had the strongest known association with excessive alcohol consumption included binge drinking, heavy drinking, liver cirrhosis mortality, alcohol-related medical admissions, and alcohol-related motor vehicle crashes, including single-vehicle night-time crashes (which are widely used to indicate the involvement of excessive drinking).[20] Less-direct measures included per capita ethanol consumption, a recognized proxy for estimating the number of heavy drinkers in a population[21]; unintentional injuries; suicide; and crime, such as homicide and aggravated assault.

Search for Evidence

The following databases were searched: Econlit, PsycINFO, Sociology Abstracts, MEDLINE, Embase, and EtOH. All years of records available on the databases were searched up to February 2008. Although the systematic search ended at this date, the review team is not aware of additional hours of sale research published since this time. (The search strategy will be available on the Community Guide website.) The reference lists of articles reviewed were also searched as well as reference lists from other systematic reviews. Government reports were considered for inclusion, but unpublished papers were not. Subject matter experts were also consulted to identify studies that might have been missed.

Assessing the Quality and Summarizing the Body of Evidence on Effectiveness

Each study that met the inclusion criteria was read by two reviewers who used standardized criteria to assess the suitability of the study design and threats to validity.[10] Uncertainties and disagreements between the reviewers were reconciled by consensus among the review team members. Classification of the study designs accords with the standards of the *Community Guide* review process and may differ from the classification reported in the original studies.

Studies were evaluated based on their design and execution. Those that collected data on exposed and control populations prospectively were classified as having the greatest design suitability. Those that collected data retrospectively or lacked a comparison group, but that conducted multiple pre- and post-measurements on their study population(s), were rated as having moderate design suitability. Finally, cross-sectional studies, those without a comparison group, and those that involved only a single pre- or post-measurement in the intervention population were considered to have the least suitable design. Quality of execution was assessed by examining potential threats to study validity, including an inadequate description of the intervention or of the study population(s), poor measurement of the exposure or outcome, failure to control for potential confounders, and a high attrition rate among study participants. Based on these criteria, studies were characterized as having good quality of execution if they had at most one threat to validity; fair execution if they had two to four threats to validity, and limited quality of execution if they had five or more threats to validity. For example, studies that used only proxy outcome measures were assigned a penalty for this threat to validity. Only studies with good or fair quality of execution were included in the body of evidence; studies with any level of design suitability were included, other than those with cross-sectional design.

Effect estimates were calculated as relative percentage change in the intervention population compared with the control population using the following formulas:

1. For studies with pre- and post-measurements and concurrent comparison groups:
   Effect estimate $=(I_{post}/I_{pre})/(C_{post}/C_{pre})-1$,
   where:
   I_{post} = last reported outcome rate or count in the intervention group after the intervention;
   I_{pre} = reported outcome rate or count in the intervention group before the intervention;
   C_{post} = last reported outcome rate or count in the comparison group after the intervention;
   C_{pre} = reported outcome rate or count in the comparison group before the intervention.
2. For studies with pre- and post-measurements but no concurrent comparison:
   Effect estimate $=(I_{post}-I_{pre})/I_{pre}$

All studies included in this review assessed the effects of *increasing* hours of sale, and the control condition was *not increasing* hours of sale. Although the analysis here accordingly assesses the effects of increasing hours, the public health intervention of interest is the control condition, (i.e., *limiting or not increasing hours of sale*). This approach rests on the assumption that increasing availability by increasing hours is likely to increase excessive consumption and related harms, and thus not increasing hours when proposed is the public health intervention. For each body of evidence, the review reports a number of *events* of policy changes in hours in a given jurisdiction, each of which may have been the subject of more than one *study* (a research investigation carried out by a single researcher or research group), each of which, in turn, may have been reported in more than one *paper* or *report*.

Results on Intervention Effectiveness

Studies of Changes of >2 Hours in Hours of Sale

Ten studies[22–31] of six events that resulted in a change of ≥2 hours in the hours of alcohol sales met the inclusion criteria. Only one study[22] was of greatest design suitability; however, the principal analysis in this study was presented graphically and did not allow the estimation of a numeric effect size. One study[23] was of moderate design suitability and eight[24–31] were of least suitable design. All studies had fair quality of execution. (A summary evidence table [Table 1][22–40] accompanies this review.)

Four of the six events studied occurred in Australia (in 1966, 1977, 1984, and 1998–2000); one in London, England (in 2005); and one in Reykjavik, Iceland (in 2005). All of the events led to increased hours of sale at on-premises alcohol outlets.

In Victoria, Australia, weekday and Saturday hours were extended from 6:00 PM to 10:00 PM in 1966. Hours allowed prior to this change were not reported. One

Table 1. Evidence of the effects of limits of alcohol hours of sale on excessive alcohol consumption and related harm

Study/design/ execution	Population/study time period	Intervention/comparison	Analysis/outcome	Reported findings	Review/effect size
Policies allowing a change of ≥2 hours—Increasing hours					
El-Maaytah (2008)[29] Design suitability: Least Pre/post, no control Quality of execution: Fair (4 limitations)	Location: University College Hospital, London, England, and Wales Dates: Intervention: November 24, 2005 Pre-intervention: November 24, 2004–April 30, 2005 Post-intervention: November 24, 2005–April 30, 2006	Intervention: Flexible opening hours: Potentially 24-hour opening, 7 days a week, dependent on special license Note: Granting of licenses subject to consideration of impact on local residents, businesses, and expert opinion Control: None	Analysis: Chi-square Outcome: ARMT (6 months before compared to 6 months after)	ARMT Pre: 1102 Post: 730	Relative % change (95% CI): −33.8% (−39.7, −27.3)
Newton (2007)[27] Design suitability: Least Pre/post, no comparison Quality of execution: Fair (3 limitations)	Location: London Dates: Intervention: November 2005 Pre-intervention: March 2005 (9:00PM–9:00AM) Post-intervention: March 2006 (9:00PM–9:00AM)	Intervention: Experimental unrestricted hours Control: None	Analysis: Mann–Whitney *U* test for differences in proportions Outcomes: Numbers and percentages of "alcohol-related" ER admissions, injuries, and hospital referrals	Significant increases in number of alcohol-related admissions, alcohol-related assault, alcohol-related injury, and alcohol-related hospital admissions	Relative % change (95% CI): Alcohol-related assault: 129.6 (46.1, 260.8) Alcohol-related injury: 193.2 (108.2, 312.8)
Babb (2007)[28] Design suitability: Least Pre/post, no comparison Quality of execution: Fair (3 limitations)	Location: London Dates: Intervention: November 2005 Pre-intervention: December 2004–November 2005 (9:00PM–9:00AM) Post-intervention: December 2005–November 2006 (9:00PM –9:00AM)	Intervention: Experimental unrestricted hours, along with fines/ penalties for service to drunk clients and children Control: None	Analysis: 30 of 43 home office police forces provide data on arrests for serious and less-serious violent crimes. Offenses not specified as alcohol-related	Moving averages calculated for nighttime arrests, 6:00PM to 5:59AM	Relative % change: Serious offenses (including homicide and manslaughter): –9.5% Less-serious offenses (with wounding): –5.4% Less-serious offenses (with wounding) in city centers and near licensed premises: –4.3% Assault without injury: –2.7% Assault without injury in city centers and near licensed premises: 3.1%
Ragnarsdottir (2002)[26] Design suitability: Least Pre/post, no comparison Quality of execution: Fair (3 limitations)	Location: "relatively small" city center, Reykjavik Dates: Intervention: July 1999–July 2000 Pre-intervention: March 1999–April 1999 (8 weekend nights) Post-intervention: March 2000–April 2000 (8 weekend nights)	Intervention: Experimental unrestricted hours Control: Unchanged hours	Analysis: Percentages; no tests of significance Outcomes: • Emergency ward admissions (not specific to city center) • Suspected drunk driving cases	For all outcomes, location not specified as city center (the location of intervention) or outside city center. Emergency ward admissions: Weekend nights: 31% increase All-day: 3% increase Weekends (all day):	Relative % change: Weekend emergency ward admissions: 20%* Accidents and other mishaps: 23%* Fighting: 34%* Suspected drunk driving: 79.3% (13.8, 182.4)

(continued on next page)

Table 1. *(continued)*

Study/design/ execution	Population/study time period	Intervention/comparison	Analysis/outcome	Reported findings	Review/effect size
	Weekend nights defined as Saturday or Sunday from 12:00 MN to 7:00AM			20% increase Weekdays: 2% decrease Reasons for admission include incidents often related to drinking: Accidents and other mishaps: 23% increase Fighting: 34% increase Non–alcohol-related admission types: No change Suspected drunk driving: 1999: 29 2000: 52	
Smith (1988)[25] Design suitability: Least Pre/post, no comparison group Quality of execution: Fair (3 limitations)	Location: Tasmania, Australia Dates: Intervention: August 10, 1977 Pre-intervention: July 1, 1971–June 30, 1977 Follow-up: October 1, 1977–September 30, 1978	Intervention: Unrestricted hours allowed throughout week. Smith reports numbers of actual hours did not change, but hours shifted to later times. Exceptions (mandatory closing): Sundays 5:00 AM–12:00NOON Sundays 8:00PM–12:00MN Good Friday Prior hotel opening hours: Monday–Saturday: 10:00 AM–10:00PM Sunday: 12:00NOON–8:00PM Control: Number of injury crash from 6:00 PM to 10:00PM	Analysis: Chi-square Outcome: Crash injury between 10:00PM and 6:00AM	Traffic injury crash: Increased between 10:00PM and 6:00AM. Although the number occurring directly after the former closing time decreased, both the proportion and the absolute number of traffic injury crash from 12:00MN to 6:00AM increased, for a total overall increase.	Relative % change (95% CI): Traffic injury crash: 10.8% (–1.5, 21.2)
Raymond (1969)[22] Design suitability: Greatest Pre/post, no comparison. Quality of execution: Fair (3 limitations)	Location: Melbourne, Victoria (Australia) Dates: Intervention: February 1, 1966 Pre-intervention: 1964–1965 Follow-up: 1966–1967 after period Note: data collection begins January 1, 1966	Intervention: Closing time extended from 6:00PM to 10:00PM Control: Sundays	Analysis: Outcomes: ● Casualty accidents ● Total accidents ○ Pedestrian accidents ○ Single-vehicle accidents ○ Multi-vehicle accidents	Summary of major findings: Total accidents: No change Hourly distribution of accidents occurring from 6:00PM to 11:00PM changed significantly: Sharp decrease from 6:00PM to 7:00PM and an increase from 10:00PM to 11:00PM.	Graphical comparison of weekdays and Saturday with hours change vs Sunday without change: No effect
Williams (1972)[23] Design suitability: Moderate Interrupted time series	Location: Victoria, Australia Dates: Intervention:	Intervention: Closing time extended from 6:00PM to 10:00PM Control: None	Analysis: Maximum likelihood estimates Outcome:	Sales increase $1.9 per quarter due to 10:00PM closing Equivalent to 12% increase	Consumption change: 12% (ns)* *CIs not calculable because of lack of data

(continued on next page)

Table 1. Evidence of the effects of limits of alcohol hours of sale on excessive alcohol consumption and related harm *(continued)*

Study/design/ execution	Population/study time period	Intervention/comparison	Analysis/outcome	Reported findings	Review/effect size
Quality of execution: Fair (2 limitations)	January 2, 1966 Pre-intervention: 1958–1966 Follow-up: 1966–1969		Consumption of alcohol in Aus$ sales per capita controlled for price of beer and consumer price index	Note: Author reports no significant effect because SEs are large	
Smith (1988)[24]	Location: Victoria, Australia Dates: Intervention: January 2, 1966	Intervention: Closing time extended from 6:00PM to 10:00PM Control: None	Injury crash change: Yearly vehicle crashes 3 years before and 1 year after the change in hours. No assessment of alcohol-relatedness of crashes	An increase of 11.5% in automobile crash injuries associated with the change in hours (not taking entire day into account)	Relative % change (95% CI): 3.6% (−16.6, 28.8)
Smith (1990)[30] Design suitability: Least Pre/post, no comparison Quality of execution: Fair (3 limitations)	Location: Victoria, Australia Dates: Intervention: (1) July 13, 1983 (2) November 1984 Pre-intervention: January 1, 1980–December 31, 1983 Follow-up (1): January 1, 1984–December 31, 1984 Follow-up (2): January 1, 1985–December 31, 1985	Intervention: (1) Two 2-hour periods allowed on Sundays between 12:00NOON and 8:00PM (2i) Full hours allowed between 12:00NOON and 8:00PM on Sunday (2ii) Monday to Saturday sales extended from 10:00PM to 12:00MN (2iii) Sunday restaurant hours increased to 12:00 NOON to 11:30PM (12:00NOON–4:00PM and 6:00PM–10:00PM) Control: None	Analysis: Chi-squares Outcome: Traffic crash injury	Injury crash during the 4 hours after 8-hour Sunday session	Relative % change (95% CI): 8.5 (2.2, 15.2)
Briscoe (2003)[31] Design suitability: Least Cross-sectional Quality of execution: Fair (3 limitations)	Location: Victoria, Australia Dates: Intervention: July 1998–June 2000	Intervention: 24-hour permit granted to some on-premises alcohol outlets	Analysis: descriptive statistics Outcomes: Number of assaults within outlets during study period	Summary of major findings: Authors claim that there is an association between 24-hour permits and high rates of assaults. However, findings appear contradictory and do not allow re-evaluation.	Inconclusive
Policies allowing a change of <2 hours					
Chikritzhs (1997)[32–35] Design suitability: Greatest Before and after design with comparison Quality of execution: Fair (3 penalties)	Location: Perth, Western Australia (WA) Dates: Data collected from July 1, 1991 to June 30, 1995 for: • Assaults Data collected from July 1, 1990 to June 30, 1996 for: • Road-block breath testing • Accidents	Intervention (1988): ETPs only (until 1:00AM instead of 12MN) Control: Hotels that served in standard hours (until 12:00MN) throughout study period (non-ETPs)	Analysis to test for ETP association: • Paired *t*-tests • Repeated measures analysis • Multiple Linear Regression Outcomes: • Monthly assault rates • Impaired driver BAL	Monthly assaults per hotel: ETP hotels: Pre: 0.121; Post: 1.87 Non-ETP hotels: Pre: 0.112; Post: 0.133 *Adjusting for alcohol sales eliminated effect of ETPs (e.g., increased consumption accounted for increased harm)	Relative % change: Monthly assaults per hotel: 30.1% Wholesale alcohol purchases: 10.5% Alcohol-related road crashes: 51.3%

(continued on next page)

Table 1. *(continued)*

Study/design/ execution	Population/study time period	Intervention/comparison	Analysis/outcome	Reported findings	Review/effect size
	• Routine police patrols		• Alcohol-related crashes • Wholesale alcohol purchase	ETP hotels: Pre: 670,403; Post: 881,048 Non-ETP hotels: Pre: 686,094; Post: 815,822 Alcohol-related road crashes: ETP hotels: Pre: 0.0781; Post: 0.0808 Non-ETP hotels: Pre: 0.0731; Post: 0.0503	
Smith (1987)[36] Design suitability: Least Before and after design, no comparison Quality of execution: Fair (3 penalties)	Location: New South Wales, Australia Dates: Intervention: Weekday/Saturday closing hours: changed from 10:00PM to 11:00PM Pre-intervention: 1976–1979 Follow-up: 1980–1981	Intervention: Hours: Weekday/Saturday evening closing hours extended from 10:00PM to 11:00PM December 1979 Sunday hours and outlet types also expanded December 1980 BAC levels lowered from 0.08% to 0.05% Control: No comparison group	Analysis: Percentage change Outcomes: Motor vehicle fatalities	Summary of major findings: Findings on this outcome not considered	Relative % change in motor vehicle fatalities: −2.7%
Knight (1980)[37] Design suitability: Least Before and after study without comparison Quality of execution: Fair (4 limitations)	Location: 4 major cities and central belt of Scotland Dates: Intervention: Hours: December 13, 1976 Pre-intervention: October–November 1976 Follow-up: March 1977	Intervention: Hours: Evening closing hours extended from 10:00PM to 11:00PM in December 1977 (Sunday licenses issued October 1977) Control: No comparison group	Analysis: Percentage changes Outcomes: Consumption and patterns of consumption	Change in consumption (in standard units) from before to after the time change: Men: −0.9 units/week Women: 0.2 units/week	Relative % change in consumption following extended hour: Men: −4.9% Women: 3.8%
Bruce (1980)[38] Design suitability: Least Before and after study with no comparison Quality of execution: Fair (2 limitations)	Location: 4 major cities and central belt of Scotland Dates: Intervention: Hours: December 13, 1976 Pre-intervention: October–November 1976 Follow-up: March 1977	Intervention: Hours: Evening closing hours extended from 10:00PM to 11:00PM in December 1977 (Sunday licenses issued October 1977) Control: No comparison group	Analysis: Percentage changes Outcomes: Beer sales in bulk barrels	Beer sales in bulk barrels Mean 1970–1976/1977 3,7856,143/40,262,000 3,264,000/366,800	Relative % change: Beer sales in bulk barrels 5.7%
De Moira (1995)[39] Duffy (1996)[40] Design suitability: Greatest	Location: England/Wales Dates: Intervention:	Intervention: Extension of opening and Sunday hours • Opening hour changed from 11:00AM to 10:00AM	Analysis: Logistic linear regression, analysis of deviance Outcomes:	Summary of major findings: Mortality: No increase in: • Liver disease and cirrhosis	Relative % changes (95% CI): Mortality from diverse alcohol-related diseases: no effect Convictions for sales to underage patrons:

(continued on next page)

Table 1. Evidence of the effects of limits of alcohol hours of sale on excessive alcohol consumption and related harm *(continued)*

Study/design/ execution	Population/study time period	Intervention/comparison	Analysis/outcome	Reported findings	Review/effect size
Prospective data collection with intervention and control populations Quality of execution: Fair (2 limitations)	August 1988 Pre-intervention: 1980–1988 Follow-up: 1988–1991	• Extra hour on Sunday (hours allowed from 12:00NOON until 10:30PM, with a mandatory break of 4 hours beginning at 3:00PM) • Drinking-up time increased from 10 to 20 minutes (weekdays only) • Off-premises sales allowed from 8:00AM Control: Scotland (positive control, having already extended hours several years previously)	• Liver disease and Cirrhosis Mortality • Pancreatitis mortality • Alcohol poisoning • Alcohol-dependent syndrome • Alcohol psychosis • Workplace absenteeism • Workplace accidents • Road accidents • Positive breath tests • Drunk driving convictions • Drunkenness offenses • Crimes of violence • Underage drinking	• Pancreatitis • Alcohol poisoning • Alcohol-dependent syndrome • Alcohol psychosis Workplace: No increase in: • Workplace absenteeism • Serious or fatal workplace accidents Increase in: • Slight workplace accidents RR Scotland: 1.34 RR E and E: 1.01 Motor vehicle: No increase in: • Drunk driving convictions • Positive breath tests • Fatal and serious road accidents Increase in: • Slight road accidents Relative % change: 3.5% Public order: No increase in: • Drunkenness offenses • Crimes of violence • Underage drinking	64.1% (21.2%, 99.0%) Purchases by minors: –62.4% (72.9%, 46.5%) Recorded violent crime: 15.5% (14.0%, 17.0%)
Vingilis (2005)[41] Design suitability: Greatest Prospective data collection with intervention and control populations Quality of execution: Fair (3 limitations)	Intervention: May 1996 Pre-intervention: 1992–1996 Follow-up: 1996–1999	Intervention: On May 1, 1996, Ontario, Canada, amended the Liquor License Act to extended closing hours for alcohol sales and service in licensed establishments from 1:00AM to 2:00AM Control: Michigan and New York states, in which similar changes did not occur	Analysis: Supposedly interrupted time series, but results not given. Graphical analyses. Outcomes: Motor vehicle fatalities, alcohol-related and all Consumption	Summary of major findings: No significant change relative to controls Declines in consumption reported	Findings: No significant change relative to controls

*CIs not calculable due to the lack of data.
ARMT, alcohol-related maxillofacial trauma; ETP, extended trading permit

study[22] compared trends in motor vehicle–related outcomes on weekdays and Saturdays before and after the hours of alcohol sales at on-premises alcohol outlets in Victoria, Australia, were extended, to the same outcomes on Sundays, when there was no change in hours. The author found that the increase in hours of sales on weekdays and Saturday did not significantly affect the number of crashes that occurred on these days. However, she observed a change in the timing of crashes corresponding to the change in the closing time of the on-premises alcohol outlets. Thus, in this study, it appeared that although the number of events may not have been affected by the change in the closing time of alcohol outlets, their timing was affected. In contrast to this study's findings, two subsequent analyses of the same event concluded that the increase in hours was associated with increases in consumption[23] and motor vehicle crash injuries.[24]

In 1984, hours available for alcohol service in Victoria were extended from 10:00PM until 12:00MN on weekdays and Saturdays and in length of time open from 4 hours to 8 hours on Sundays (a day on which alcohol sales had been previously allowed). Information on hours prior to the weekday and Saturday extension is not given. A study of this event[30] found an increase in motor vehicle crash injuries associated with these increases in hours.

Between July 1998 and June 2000, Victoria granted 24-hour permits to some on-premises alcohol outlets. A cross-sectional study comparing rates of assaults in outlets granted and not granted 24-hour permits is inconclusive.[31] Although authors claim that higher rates of assault are associated with 24-hour facilities, their statements describing results are inconsistent, and the authors do not provide data to allow re-evaluation.

In Tasmania (Australia), licensed premises were allowed to stay open until any hour in 1977. Prior Monday–Saturday opening hours were 10:00AM–10:00PM; Sunday hours, 12NOON– 8:00PM. The assumption by policymakers underlying unrestricted closing times was that possibly intoxicated clients would not be exiting the facilities at the same time, potentially decreasing risks, because different outlets would choose different closing hours. A study of this event[25] found an increase in motor vehicle crash injuries associated with these increases in hours.

In Reykjavik, licensed premises were allowed to stay open until any hour in the year 1999 on an experimental basis. Prior closing requirements were 11:30 PM on weekdays and 2:00 AM on weekends. Researchers found increases in emergency room admissions, injuries, fighting, and suspected driving while intoxicated.[26]

Finally, the United Kingdom's Licensing Act of 2003 allowed sales of alcoholic beverages 24 hours a day in England and Wales, beginning in November 2005, subject to local licensing requirements. Three studies assessing the impact of this increase in hours of sale produced mixed results.[27–29] Two studies[28,29] found a relative decrease in harms (violent criminal offenses and alcohol-related maxillofacial trauma, respectively), whereas a third study[27] found a relative increase in harms (alcohol-related assault and injury) subsequent to this increase in hours of sale.

Among the ten studies in this body of evidence,[22–31] two studies[28,29] found that an increase of ≥2 hours in the hours of sale led to decreased alcohol-related harms (i.e., injury and serious violent crime), and six studies[23–27,30] found an increase in alcohol-related harms relative to the period before the increase in hours of sale took place (Figure 2). The study by Raymond[22] found no effect. One study[23] found a nonsignificant increase in alcohol consumption associated with the increase in hours in Victoria, Australia, in 1966.

Information on the requirement that premises seek permits prior to expanding hours may not have been complete in the studies reviewed. To the extent that stated permit requirements accurately reflect the expansion process, there appears to be no systematic effect of permitting. Although the harmful effects of permitted expansions appear to be larger than those in which permits were not required (Figure 2) there were also effects in the opposite direction for studies of permitted settings.

Studies of Changes of <2 Hours in Hours of Sale

Six studies of five events (reported in ten papers[32–41]) that resulted in a change of <2 hours of sale met the inclusion criteria. All studies were of on-premises alcohol outlets. Three studies (seven papers[32–35,39–41]) were of greatest design suitability, three[36–38] were of least suitable design; all were of fair quality of execution. One study (two papers[39,40]) of the extension of opening hours in England and Wales in 1988 did not allow the calculation of effects for several outcomes, but it reported small and inconsistent results on multiple alcohol-related outcomes. One[41] provides graphics and report using interrupted time series but does not report numeric results.

In 1993, Perth, Australia allowed on-premises outlets to extend their closing time from 12:00MN to 1:00AM.[32–35] Findings were inconsistent, with a reported increase of alcohol wholesale but a decline in drunk driving and an increase in assaults and in alcohol-related crashes. None of these findings was significant.

In December 1979, the state of New South Wales in Australia expanded on-premises alcohol outlet closing hours from 10:00PM to 11:00PM, at the same time expanding Sunday hours and outlet settings. A study of these events[36] proposed using the weekdays as the control in an assessment of the effects of increased Sunday sales on

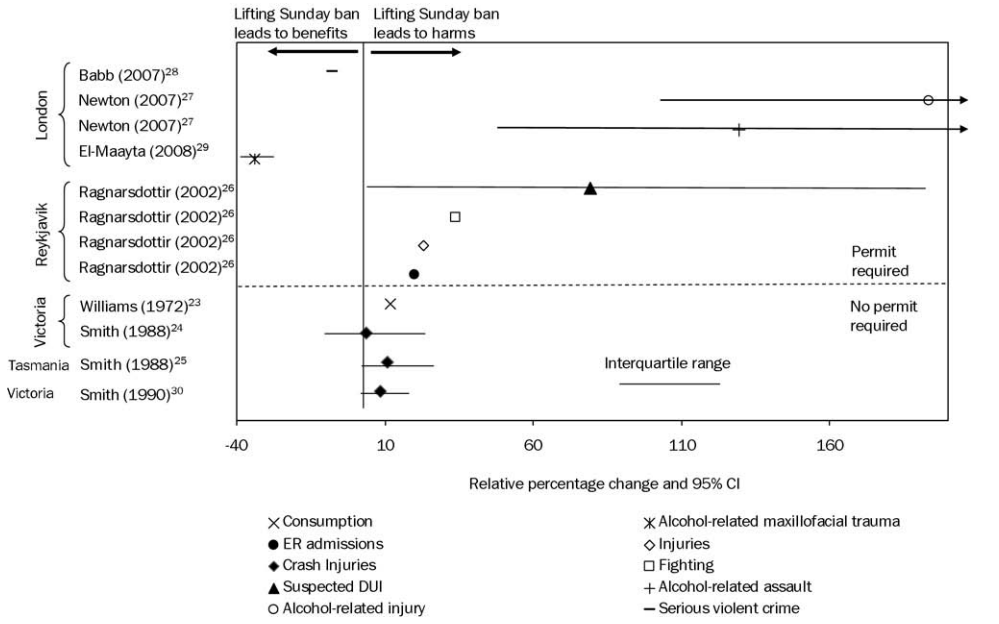


Figure 2. Relative percentage change in diverse outcomes associated with increases of ≥2 hours
DUI, driving under the influence

motor vehicle fatalities. However, this comparison is biased toward a null effect, given the change in weekday hours. A comparison of weekday fatalities before and after the weekday expansion indicates a reduction of 2.7% in motor vehicle fatalities over the study period associated with the weekday increase of 1 hour in closing time. However, this outcome may be confounded by a reduction from 0.08% to 0.05% in maximum legal blood alcohol levels in December 1980, which would have been expected to deter drunk driving and reduce motor vehicle injuries.

In 1976, Scotland allowed on-premises outlets to extend their closing time from 10:00PM to 11:00PM.[37,38] Reported changes were small and not consistent in direction. Knight found increased consumption for women and decreased consumption for men, and Bruce reported a small increase in the per capita consumption of beer.

In 1988, England and Wales extended the closing hours at on-premises outlets from 10:30PM to 11:00PM and moved the opening time from 11:00AM to 10:00AM.[39,40] The outcomes, including mortality from liver disease and cirrhosis, pancreatitis, alcohol poisoning, "alcohol-dependent syndrome," alcohol psychosis, workplace absenteeism and injury, and various motor vehicle–related outcomes) assessed in these studies were heterogeneous and included the seemingly contradictory findings that in comparison with changes in the control setting (Scotland), convictions for sales to underage patrons increased by 64.1% (95% CI=21.2%, 99.0%), whereas sales to minors fell substantially. Another finding was an increase in recorded violent crime of 15.5% (95% CI= 14.0%, 17.0%). (See Table 1.)

Finally, in 1996, Ontario Province extended closing hours in on-premises alcohol outlets from 1:00AM to 2:00AM. A study[41] of this event used graphics and interrupted time series to assess the effects of this change on all and alcohol-related fatal motor vehicle crashes. Changes in Ontario were compared with changes in Michigan and New York, neither of which changed hours of sale during the same period. The study also assessed changes in the sales of beer, wine, and spirits in Ontario from the period before to the period following the policy change. Numeric results are not reported. Beer consumption declined over the study period, whereas the consumption of wine and spirits declined in the early 1990s and then increased in the later 1990s. The authors conclude that changes in motor vehicle outcomes are "minimal." Their graphics suggest a shift of the timing of alcohol-related fatalities to later hours following the extension of hours of sale.

This small body of evidence indicates no consistent effects of changes of <2 hours on alcohol-related outcomes. Four events of increases in hours of sale were studied. Only one study of increased hours of sale in Perth, Australia, reported substantial increases in wholesale alcohol purchases, assaults, and motor vehicle crashes. Two studies (of events in England and Wales and in Ontario, Canada) did not provide numeric results but reported small and inconsistent changes in alcohol-related outcomes including alcohol consumption, multiple alcohol-related causes of mortality, and motor vehicle crashes. Two studies of increased hours of sale in Scotland also reported small and inconsistent changes in alcohol sales and consumption.

Again, information on the requirement that premises seek permits prior to expanding hours may not have been complete in the studies reviewed. To the extent that stated permit requirements accurately reflect the expansion

process, there appears to be no systematic effect of permitting (Figure 3).

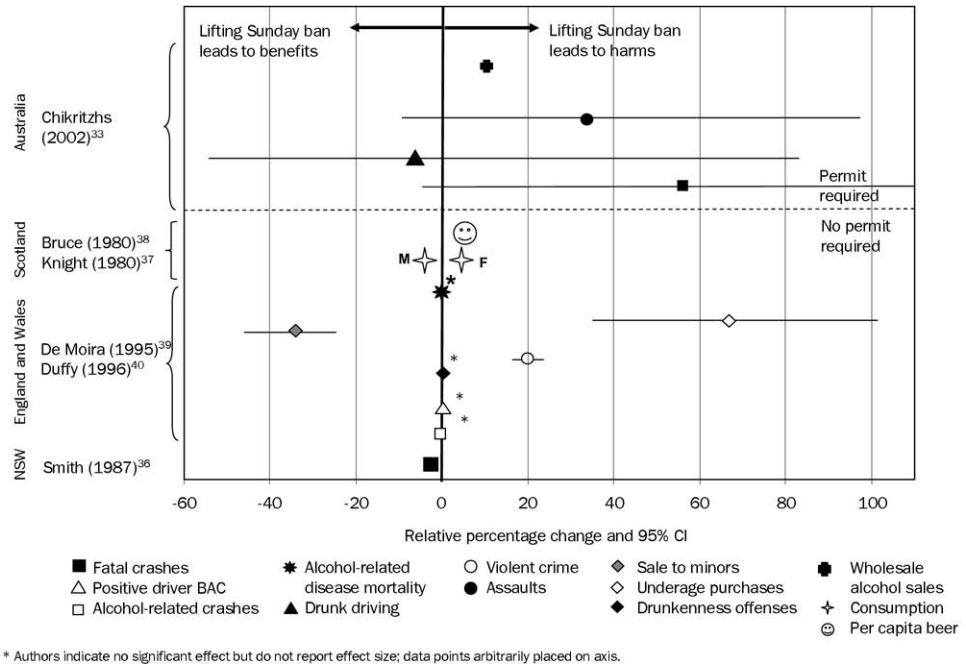


Figure 3. Relative percentage change in diverse outcomes associated with increases of <2 hours
NSW, New South Wales

Applicability

The studies in this review were conducted in a variety of settings outside the U.S. and during a wide range of time periods. Nonetheless, the association between restrictions on the hours when alcohol may be sold and alcohol-related harms was consistent across most geographic locations (all in high-income countries) and time periods, and the findings of this review are likely to be relevant for considering the potential impact of modifying the number of hours when alcohol may be sold in the U.S.

Other Harms and Benefits

Maintaining hours of sale may sustain quality of life in communities by controlling alcohol availability, excessive alcohol consumption, and health and social harms resulting from excessive alcohol use (e.g., public drunkenness); evidence of effects on quality of life were not provided by the studies reviewed. Although it is possible that crimes such as illicit alcohol sales may increase in localities where the hours of sale are limited, no evidence of such effects was found in any of the studies evaluated. One study[26] noted increased workload among law enforcement personnel associated with expanded hours of sale.

Barriers

The maintenance and reduction in the number of hours when alcohol may be sold may affect overall alcohol sales and may thus be opposed by commercial interests involved in manufacture, distribution, and sale of alcoholic beverages. The alcohol industry has generally supported policies that remove restrictions on the access to alcohol.[42]

State pre-emption laws (i.e., state laws that prevent the implementation and enforcement of local policies more restrictive than statewide regulations) can also undermine efforts by local governments to regulate hours of sale.[6] Indeed, the elimination of pre-emption laws related to the sale of tobacco products is one of the health promotion objectives in *Healthy People 2010*.[5] However, there is no similar objective in *Healthy People 2010* related to the local sale of alcoholic beverages.

Economics

No studies were identified that assessed the economic impact of reducing the number of hours when alcohol may be sold. No study was found that specifically estimated the magnitude of commercial losses in sales and tax revenues because of a policy of restricting hours of alcohol sales.

Summary

This review found that increasing the hours when alcohol may be sold by ≥2 hours increased alcohol-related harms. Evidence supporting this conclusion was based on studies conducted in on-premises settings outside the U.S. According to *Community Guide* rules of evidence, these findings provided sufficient evidence for the effectiveness of maintaining limits on hours of sale for the reduction of alcohol-related harms when efforts are made to increase hours by ≥2.[10] Because no qualifying study assessed the effects of reducing hours of sale, the only direct inference that can be made is that reducing hours of sale by ≥2 is likely to avert alcohol-related harms. How-

ever, it may also be reasonable to expect that reducing hours of sale would also reduce alcohol-related harms.

Because there was no consistent effect on excessive alcohol consumption or related harms of increasing hours of sales by <2 hours, according to *Community Guide* rules of evidence, there was insufficient evidence that this intervention had a meaningful effect.[10] Insufficient evidence means that it is not possible to determine from the available evidence whether this policy change had a meaningful effect.

Research Gaps

All existing research on hours of sale to date has been conducted in nations other than the U.S. It would be useful to have studies of changes in hours of sale in U.S. settings to confirm results from other settings. In addition, all research thus far has assessed the effects of *increasing* hours of sale. Although it may be a less-frequent event, evaluating the effects of *reducing* hours of sale for preventing excessive alcohol consumption and related harms would be useful. Evidence on changes in hours of sale of <2 hours is currently insufficient because of inconsistent findings. Thus, when such changes occur, it may be worthwhile to assess the effects of smaller changes in hours of sale on excessive alcohol consumption and related harms to improve our understanding of the "dose–response" and "threshold" relationships between changes in hours of sale and public health outcomes.

Additional research is also needed to more fully assess the costs and benefits of restricting the number of hours when alcohol is sold. From a societal perspective, economic elements should include intervention costs; loss in sales, tax revenues, and employment; reductions in fatal and nonfatal injuries, crime, and violence; gains in safety and public order; and averted loss of household and workplace productivity.

Finally, no studies were found that assessed the effects of changes in hours of sale in off-premises settings. Although consumers at off-premises settings are less likely to be directly affected by the effects of excessive consumption at the place of purchase, it is nevertheless possible that changes in availability in these settings may also affect alcohol-related harms. This issue merits investigation.

Discussion

Based on a systematic review of qualifying studies, this review confirms the findings of previous reviews and adds details regarding a possible dose or threshold effect. Evidence of the effects of changes in hours of sale of <2 hours was insufficient to determine effectiveness because of inconsistency among findings in the body of evidence, leaving unanswered the question of the effects of small increases in hours of sale. Data are not sufficient to allow systematic assessment of the relative percentage increase in hours (over a baseline) or the placement of the hours within the day.

All of the studies included in this review assessed the effects of increasing hours of sale at on-premises outlets, consistent with the international trend toward expanding the availability of alcoholic beverages. Further scientific evidence is needed to fully assess the symmetry between the effects of maintaining existing limits on the hours of sale compared with reducing hours of sale.

The only available evidence of the effects of reducing hours of sale was from a study in Brazil,[43] which did not qualify for inclusion in the review because Brazil is not a high-income nation, and, in general, studies of alcohol consumption from middle- and lower-income nations are thought not to be directly applicable to the contemporary U.S. context. In 1999, the city of Diadema had very high homicide rates; 65% of these were alcohol-related. Most of the homicides occurred between 11:00PM and 6:00AM. Diadema law allowed 24-hour opening of alcohol outlets. In July 2002, a new city law required bars to close at 11:00PM. From 2002 to 2005, homicide rates in the city declined by 44% (95% CI=27%, 61%), controlling for mortality trends. During this time period, there was also a 17% decline in assaults against women (the only additional outcome assessed); this finding, however, was not significant.

In addition to the lack of studies that assessed the effect of stricter limits on the hours when alcohol may be sold, the body of qualifying studies in this review had several other limitations. First, some studies did not directly assess the impact of relaxing restrictions on the hours of sales on excessive alcohol consumption and alcohol-related harms, but rather relied on proxy measures of these effect outcomes (e.g., criminal arrest rates). Second, nearly all of the studies relied on population-based data from public health surveillance systems that did not capture information on alcohol control policies. As a result, many of these studies were unable to control for some potential confounding factors. However, these studies generally assessed changes in the same geographic area before and after the implementation of changes in hours of sale over a fairly short time period. Other contextual factors that could also influence alcohol sales and consumption (e.g., changes in alcohol excise taxes) at the country, state, or community levels were likely to have remained fairly constant during the study periods, allowing for a valid assessment of the impact of changing hours of sale, independent of other factors, on alcohol-related harms.

The findings in this review support the potential value of allowing local communities to maintain restrictions on hours of sale. If further research supports the effectiveness of local restrictions on hours of sale, it would also argue for eliminating state pre-emption laws that prohibit local governments from enacting alcohol control policies more restrictive than those that exist statewide.

We acknowledge the support and contributions of Steven Wing of the Substance Abuse and Mental Health Services Administration (SAMHSA) in discussions of this and other reviews.

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 Mortal Wkly Rep 2004;53(37):866–70.
2. National Center for Health Statistics. Health, U.S., 2005 with chartbook on trends in the health of America. Hyattsville MD: NCHS, 2005. Report No.: 1232.
3. 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.
4. 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. Rockville MD: NIAAA, 2000. Report No.: 98-4327.
5. USDHHS. Healthy People 2010. www.healthypeople.gov/.
6. Mosher J. Alcohol issues policy briefing paper: the perils of preemption. Chicago: American Medical Association, 2001.
7. Gorovitz E, Mosher J, Pertschuk M. Preemption or prevention? Lessons from efforts to control firearms, alcohol, and tobacco. J Public Health Policy 1998;19(1):36–50.
8. Pub. L. No. 83-277, 67 Stat. 586 (1983).1983.
9. Wikipedia.org. Alcohol laws of the U.S. by state. en.wikipedia.org/wiki/Alcohol_laws_of_the_United_States_by_state.
10. 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.
11. Grover PL, Bozzo R. Preventing problems related to alcohol availability: environmental approaches. DHHS, SAMHSA, CSAP, 1999.
12. Stockwell T, Gruenewald P. Controls on the physical availability of alcohol. In: Heather N, Peters TJ, Stockwell T, eds. International handbook of alcohol dependence and problems. Chichester, UK: Wiley, 2004:699–719.
13. Smith DI. Effectiveness of restrictions on availability as a means of preventing alcohol-related problems. Contemp Drug Prob 1988;627–84.
14. Popova S, Giesbrect N, Bekmuradov D, Patra J. Hours and days of sale and density of alcohol outlets: impacts on alcohol consumption and damage: a systematic review. Alcohol Alcohol 2009;44:500–16.
15. Babor TF, Caetano R, Casswell S, et al. Alcohol: no ordinary commodity-research and public policy, 2nd edition. Oxford, UK: Oxford University Press, 2010.
16. WHO. Evidence for the effectiveness and cost-effectiveness of interventions to reduce alcohol-related harm. Copenhagen, Denmark: WHO Regional Office for Europe, 2009. www.euro.who.int/__data/assets/pdf_file/0020/43319/E92823.pdf.
17. Strategic Task Force on Alcohol. Strategic Task Force on Alcohol—second report. Ireland: Health Promotion Unit, Department of Health and Children, 2004.
18. Lipsey MW, Wilson DB, Cohen MA, Derzon JH. Is there a causal relationship between alcohol use and violence? In: Galanter M, ed. Recent developments in alcoholism: volume 13, alcohol and violence. New York: Plenum Press, 1997:245–82.
19. World Bank. World development indicators 2006. devdata.worldbank.org/wdi2006/contents/cover.htm. 2006.
20. Gruenewald PJ, Millar AB, Treno AJ, Yang Z, Ponicki WR, Roeper P. The geography of availability and driving after drinking. Addiction 1996;91(7):967–83.
21. Cook PJ, Skog OJ. Alcool, alcoolisme, alcoolisation—comment. Alcohol Health Res World 1995;19(1):30–1.
22. Raymond A. Ten o'clock closing—the effect of the change in hotel bar closing time on road accidents in the metropolitan area of Victoria. Aust Road Res 1969;3(10):3–17.
23. Williams RA. Changes in trading hours: ten o'clock closing and consumption of alcohol in Victoria. Econ Record 1972;48:123–7.
24. Smith DI. Effect on casualty traffic accidents of the introduction of 10 p.m. Monday to Saturday hotel closing in Victoria. Aust Drug Alcohol Rev 1988;7:163–6.
25. Smith DI. Effect on traffic accidents of introducing flexible hotel trading hours in Tasmania, Australia. Br J Addict 1988;83:219–22.
26. Ragnarsdottir T, Kjartansdottir A, Davidsdottier S. Effect of extended alcohol serving hours in Reykjavik, Iceland. In: Room R, ed. The effects of Nordic alcohol policies. Helsinki, Finland: Nordic Council for Alcohol and Drug Research, 2002:145–54.
27. Newton A, Sarker SJ, Pahal GS, van den Bergh E, Young C. Impact of the new UK licensing law on emergency hospital attendances: a cohort study. Emerg Med J 2007;24:532–4.
28. Babb P. Violent crime, disorder and criminal damage since the introduction of the Licensing Act 2003. Home Office Online Report 16/07. 2007.
29. El-Maaytah M, Smith S, Jerjes W, et al. The effect of the new "24 hour alcohol licensing law" on the incidence of facial trauma in London. Br J Oral Maxillofac Surg 2008;46(6):460–3.
30. Smith DI. Effect on casualty traffic accidents of changing Sunday alcohol sales legislation in Victoria, Australia. J Drug Issues 1990;20(3):417–26.
31. Briscoe S, Donnelly N. Problematic licensed premises for assault in inner Sydney, Newcastle and Wollongong. Aust N Z J Criminol 2003;36(1):18–33.
32. Chikritzhs T, Stockwell T, Masters L. Evaluation of the public health and safety impact of extended trading permits for Perth hotels and nightclubs. Perth: National Drug Institute, 1997.
33. Chikritzhs T, Stockwell T. The impact of later trading hours for Australian public houses (hotels) on levels of violence. J Stud Alcohol 2002;63(5):591–9.
34. Chikritzhs T, Stockwell T. The impact of later trading hours for hotels on levels of impaired driver road crashes and driver breath alcohol levels. Addiction 2006;101(9):1254–64.
35. Chikritzhs T, Stockwell T. The impact of later trading hours for hotels (public houses) on breath alcohol levels of apprehended impaired drivers. Addiction 2007;102(10):1609–17.
36. Smith DI. Effect on traffic accidents of introducing Sunday hotel sales in New South Wales, Australia. Contemp Drug Prob 1987;14:279–94.
37. Knight I, Wilson P. Scottish licensing laws. London: Office of Population Censuses and Surveys, Social Survey Division, 1980.
38. Bruce D. Changes in Scottish drinking habits and behaviour following the extension of permitted evening opening hours. Health Bull 1980;38(3):133–7.
39. De Moira ACP, Duffy JC. Changes in licensing law in England and Wales and alcohol-related mortality. Addiction Res 1995;3(2):151–64.

40. Duffy JC, De Moira ACP. Changes in licensing law in England and Wales and indicators of alcohol-related problems. Addiction Res 1996;4(3):245–71.
41. Vingilis E, McLeod AI, Seeley J, Mann RE, Beirness D, Compton CP. Road safety impact of extended drinking hours in Ontario. Accid Anal Prev 2005;37(3):549–56.
42. Giesbrecht N. Roles of commercial interests in alcohol policies: recent developments in North America. Addiction 2000;95(4): S581–95.
43. Duailibi S, Ponicki W, Grube J, Pinsky I, Laranjeira R, Raw M. The effect of restricting opening hours on alcohol-related violence. Am J Public Health 2007;97(12):2276–80.

Curr HIV/AIDS Rep (2010) 7:226–233
DOI 10.1007/s11904-010-0060-6

Alcohol and HIV Disease Progression: Weighing the Evidence

Judith A. Hahn · Jeffrey H. Samet

Published online: 3 September 2010

Abstract Heavy alcohol use is commonplace among HIV-infected individuals; however, the extent that alcohol use adversely impacts HIV disease progression has not been fully elucidated. Fairly strong evidence suggests that heavy alcohol consumption results in behavioral and biological processes that likely increase HIV disease progression, and experimental evidence of the biological effect of heavy alcohol on simian immunodeficiency virus in macaques is quite suggestive. However, several observational studies of the effect of heavy alcohol consumption on HIV progression conducted in the 1990s found no association of heavy alcohol consumption with time to AIDS diagnosis, while some more recent studies showed associations of heavy alcohol consumption with declines of CD4 cell counts and nonsuppression of HIV viral load. We discuss several plausible biological and behavioral mechanisms by which alcohol may cause HIV disease progression, evidence from prospective observational human studies, and suggest future research to further illuminate this important issue.

J. A. Hahn (✉)
Division of Infectious Diseases, Department of Medicine,
San Francisco General Hospital, University of California,
San Francisco, Box 0811 UCSF,
San Francisco, CA 94143-0811, USA
e-mail: judy.hahn@ucsf.edu

J. H. Samet
Department of Medicine, Section of General Internal Medicine,
Clinical Addiction Research and Education (CARE) Unit,
Boston University School of Medicine/Boston Medical Center,
Boston, MA 02118, USA

J. H. Samet
Department of Social and Behavioral Sciences,
Boston University School of Public Health,
Boston, MA 02118, USA

Keywords Alcohol · HIV disease progression · CD4 · HIV viral load · Adherence · Prospective studies · Nutrition · Immune activation · Bacterial translocation

Introduction

Heavy alcohol use is commonplace among HIV-infected individuals and its impact upon behaviors and the course of disease has been examined over the past two decades [1–3]. However, the extent that alcohol use results in deleterious effects on the progression of HIV disease has not been fully elucidated. Although alcohol may directly affect HIV disease progression in multiple ways, obtaining data to document its overall impact and contributions associated with specific mechanisms is difficult. Alcohol consumption has direct effects on several aspects of the immune system; yet, experimental studies to assess more directly its impact on HIV disease have been limited to the related model of simian immunodeficiency virus (SIV) in primates. In humans, observational studies may be hampered by measurement error and confounding. Incomplete assessment of behavioral (eg, other substance use, medication adherence, nutritional deficiencies) and psychosocial (eg, depressive symptoms) factors associated with HIV disease progression impedes rigorous determination of alcohol's direct effects. In addition, traditional analytic methods may fail to account for the potential feedback loop between alcohol consumption and health status (ie, that alcohol consumption tends to decrease as health declines). In this article we review the empirical studies and the major mechanisms by which alcohol may affect HIV disease progression (Fig. 1) based on the literature available as of early 2010.

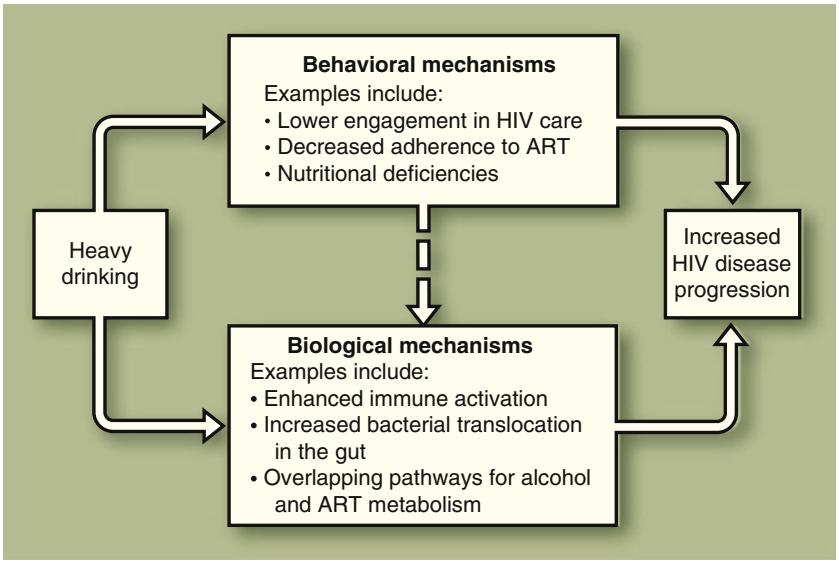


Fig. 1 Potential mechanisms of HIV disease progression. ART—antiretroviral therapy

HIV Disease Progression: Definitions

Progression of HIV infection in humans or SIV infection in the HIV primate model has been defined most typically in terms of viral load, CD4 cell count, AIDS-defining clinical end points (eg, opportunistic infections), and mortality. More recent evidence of the association of the level of immune activation with clinical outcomes is increasingly recognized; however, few data exist at present to consider this marker of HIV disease progression. In this literature review we sought published research that provided measurement of alcohol consumption as well as one of the following measures of HIV disease progression: HIV viral load, CD4 cell count, opportunistic infections, or death.

Alcohol Use: Definitions

Alcohol use can be defined by the amount consumed (eg, at-risk, heavy) or by the consequences of its use (eg, abuse, dependence). The consumption threshold for at-risk use as defined by the National Institute on Alcohol Abuse and Alcoholism (NIAAA) is >4 drinks on an occasion or >14 drinks in a week for men, >3 and >7 respectively, for women [4]. This level of drinking has been used to define "heavy" drinking for those with HIV infection [5••], although its use was chosen based on NIAAA's use of the at-risk levels in the general population and was not specific for HIV-infected individuals. The category of "moderate" alcohol consumption is used for one whose alcohol use falls between abstinence (ie, no use of alcohol) and heavy use.

Another dimension to recognize when defining alcohol use is the variable nature of its use, in that for almost half of an HIV-infected cohort of persons with a history of alcohol problems, use was observed to increase or decline over a median period of 3.4 years [6]. Studies that only measure drinking at baseline fail to account for the dynamic nature of alcohol consumption, which may adversely affect study results and conclusions.

Biological Basis for Alcohol Affecting HIV Disease Progression

The Effects of Alcohol on the Immune System

Chronic alcohol consumption has been shown to be associated with increased susceptibility to infectious diseases (eg, tuberculosis and bacterial pneumonia), increased severity of diseases (eg, viral hepatitis), and increased risk of cancers (eg, hepatocellular carcinoma) [7]. These illnesses may also be accelerated by vitamin deficiencies, malnutrition, or other substance use. Chronic alcohol consumption can lead to liver disease and cirrhosis, which can impact immunocompetence. Recent animal and human studies have shown that alcohol consumption has deleterious effects on both the innate and the acquired immune responses. Impaired innate immune responses may cause susceptibility to infection, while impaired acquired immune responses such as impaired B lymphocyte function, altered cytokine balance, and chronic T-cell activation may accelerate disease progression, including that caused by HIV. In addition, alcohol may play a role in translocation of bacteria and bacterial products from the gut to cause HIV immune activation, resulting in increased HIV disease progression [8].

Alcohol and Nutrition

Micro-nutrient (ie, vitamin and mineral) deficiencies, including selenium and vitamins D, A, B-12, and E, zinc,

and iron, have been associated with more rapid HIV disease progression [9–11]. At the same time, antiretroviral therapy (ART) is thought to decrease some and restore other micronutrient levels [12, 13]. Overall nutritional status has been associated with HIV progression in several prospective studies [14–16], while there is also an effect of HIV infection on nutritional status [17]. Alcohol consumption is associated with nutritional deficiencies, due to a high percentage of caloric intake from alcohol, decreased absorption of nutrients, and interference with the metabolism of nutrients [18]. Therefore, it is plausible that nutritional deficiency is a mechanism by which alcohol might result in more rapid HIV disease progression. The link between alcohol use, decreased nutrition, and immune markers has been demonstrated experimentally in the SIV model [19]; however, it is not clear for HIV infection in humans.

Alcohol and ART Effectiveness

Some ART, including non-nucleoside reverse transcriptase inhibitors and protease inhibitors, is metabolized by the human cytochrome P450 system. There is evidence that alcohol may impact the metabolism of these medications by two different mechanisms, enzymatic induction [20], associated with chronic alcohol use, and enzymatic inhibition due to competition of ethanol for various cytochrome P450 isozymes, associated with acute ethanol use [21]. Because numerous drugs are metabolized by the P450 pathway, chronic alcohol users may be at higher risk for drug toxicities and ineffective therapy due to inadequate plasma drug concentration. Chronic alcohol consumption may also alter drug protein binding. We examine the evidence as to whether heavy alcohol consumption reduces the effectiveness of ART as measured by failure to achieve viral suppression, even with good adherence.

Behavioral/Psychosocial Basis for Alcohol Affecting HIV Disease Progression

Several behavioral and psychosocial factors that are associated with heavy alcohol consumption are also associated with HIV disease progression, and therefore should be considered when conducting multivariate modeling. We summarize these associations in the following sections.

Access to and Retention in HIV Care and Receipt of ART

Researchers in the United States and internationally have found that heavy drinkers are less likely than others to be receiving ART [22•, 23, 24••]. This may be a consequence of barriers to consistent medical treatment or due to physicians' impressions that heavy drinkers are unable to competently use ART. Indeed, heavy alcohol use was independently associated with lower retention in care among indigent patients [25]. Because early and consistent HIV care and receipt of opportunistic infection (OI) prophylaxis and ART are key factors in slowing the progression of HIV [26], these findings suggest one mechanism by which heavy alcohol consumption may accelerate HIV disease progression.

ART Adherence

Alcohol consumption has been consistently associated with poorer ART adherence. A recent meta-analysis found that those who used alcohol or drank relatively more were 50%–60% as likely to be classified as adherent compared with those who abstained or drank relatively less [27•]. Some studies have found a dose-response relationship between alcohol consumption and ART adherence [24••, 28], in addition to a temporal association between drinking episodes and missed doses [28]. As ART adherence is a known predictor of HIV outcome, this strongly suggests that decreased adherence is one mechanism by which alcohol may increase HIV disease progression. These findings elevate the importance of attempting to control for ART adherence in examining the biological impact of alcohol on HIV disease progression *apart* from its impact on adherence. However, adherence is challenging to accurately measure and therefore may pose problems even if included in statistical models. Alternatively and likely more effectively, studies of persons infected with HIV who are not yet on ART may reveal key insights about this important issue.

Alcohol and Comorbidities

Comorbidities may complicate the issue of alcohol's impact on HIV disease progression and make it difficult to determine the association. For example, depression, which can be exacerbated by the effects of alcohol [29], has been shown to reduce adherence to ART [30], and depression, stress, and trauma may have worsening effects on HIV progression [30, 31].

Illicit substance use is strongly associated with heavy alcohol consumption, especially in minority and inner-city populations. Several studies have shown associations between illicit substance use, especially crack cocaine, and HIV progression [32•, 33]. Substance use other than alcohol has also been associated with lower rates of receipt of ART [24••]. Hence, controlling for other non-alcohol substance use is important in seeking an understanding of the impact of alcohol consumption on HIV disease progression.

Evidence of a Causal Association between Alcohol Use and HIV Disease Progression

Experimental Studies

The direct impact of alcohol consumption on HIV disease progression has been studied using animal models. The strengths of these studies include the absence of ART adherence as a possible mediator of disease progression, and the capacity to control the quantity of alcohol consumed as well as behavioral factors (eg, nutrition), which are not easily accounted for in human studies. Evidence from studies in macaques suggests that heavy alcohol consumption has consequences for increased SIV progression. Experimental administration of the equivalent of heavy doses of alcohol as compared to sucrose to macaques infected with SIV was associated with early plasma CD4 cell loss in some studies [34, 35] but not others [36, 37], while several studies found greater SIV viral load at various times post-infection [19, 34, 36, 37]. Bagby et al. [36] found a significantly more rapid onset of end-stage disease in eight alcohol-administered macaques compared with eight controls. Higher viral load in alcohol-exposed macaques was associated with a higher percentage of SIV target cells (CD4) in the gut coupled with lower percentages of CD8 cells, creating a blunted mucosal immune response in early infection in one study [37]. The alcohol-exposed group consumed significantly fewer calories than the controls in another study [19]. Taken together, these findings suggest a biologically deleterious effect of heavy alcohol administration on disease progression in SIV-infected primates.

Human Observational Studies: Pre-Highly Active Antiretroviral Therapy (HAART)

A number of clinical studies assessed alcohol use in cross-sectional and prospective analyses. Associations between HIV outcomes such as CD4 cell counts and HIV viral load in cross-sectional studies may reflect differences in the time of study entry by alcohol consumption category, therefore we will focus in this review on prospective studies. Several prospective studies were published using data collected in the pre-HAART era, as described below.

An analysis conducted in the Multicenter AIDS Cohort Study (MACS) included 1706 HIV-infected men, and examined alcohol consumption by average number of drinks per day, ranging from zero to greater than two, the latter meeting the threshold for "heavy" drinking in men. There was no association between drinks per day both at enrollment and prospectively and the development of AIDS [38]. A later analysis of the MACS determined that decreasing alcohol consumption (ie, having a significantly negative slope in the average number of drinks per week in the prior 6 months) was associated with developing AIDS-related conditions, suggesting a decrease in alcohol consumption as HIV progressed [39]. Studies conducted in men in the Netherlands [40] and Norway [41] found no association between the number of drinks per day in the prior 6 months or daily drinking, respectively, with the development of AIDS. A study of vitamin deficiencies (n= 312) reported that in a multivariate model that included age, HIV-related symptoms, baseline CD4 cell count, and several vitamin concentrations, frequent alcohol consumption (>2 times/week) at study baseline was associated with increased time to CD4 cell count declines to 200 cells/mm^3 and time to AIDS [42]. A study of 403 persons seroconverting during the Tricontinental Seroconverter Study found that any alcohol use during the first three quarters of the follow-up period, limited to avoid the feedback loop between symptoms and alcohol consumption, was not associated with time to AIDS or death [43]. The selection of "any alcohol use" as a main independent variable, as in the latter study, is a coarse measure to assess alcohol's impact on HIV disease progression. All of the preceding studies were conducted among men who had sex with men.

Two early studies were conducted among injecting drug users. One study among 496 HIV-infected methadone maintenance patients found no association between daily alcohol consumption in the prior month and time to AIDS or death in a time-dependent multivariate model that included age, sex, CD4 cell count, zidovudine use, having two or more symptoms, and crack cocaine use, while crack cocaine use was independently associated with progression to AIDS [44]. A study conducted among 188 injection drug users found that very heavy alcohol consumption (>21 drinks per week) at baseline was associated with increased %CD8 cells 2–5 years after seroconversion; no impact on CD4 cell count or %CD4 was found [45].

The only study of the issue from a developing country was a cohort study of 105 HIV clinic patients who were not on ART, conducted in Zimbabwe. This study showed no association between any alcohol consumption at baseline and successive CD4 cell count and HIV viral loads over a period of 6 months; however, follow-up was quite limited and changes in these outcomes were not examined [46].

In summary, in the pre-HAART era, no association of alcohol use with more rapid HIV disease progression was identified; however, some studies' measurement of alcohol consumption was limited in detail or only measured at baseline, and the studies largely examined men.

Human Observational Studies: Post-HAART Studies

Studies conducted after the introduction of HAART differed from the earlier studies in that most used CD4 cell

count and/or HIV viral load as the study outcomes, an understandable strategy given the reduction in OI and death in the HAART era. Notably, these studies either stratified by or controlled for ART use, and the measurement of alcohol use was more detailed than in the previous studies.

Chander et al. [24••] reported observations among 1711 persons enrolled in an urban HIV clinic cohort from 1998–2003. They found that heavy alcohol use in the prior 6 months alone and combined with injection drug use was associated with decreased viral load suppression after adjusting for age, race, nadir CD4 cell count, and years on ART. Controlling for self-reported adherence in these analyses attenuated the effect of heavy drinking, providing evidence for the causal chain between alcohol and HIV outcomes via ART adherence. Because the attenuation of the effect from a 24% to a 14% reduction in odds of viral suppression was accompanied by somewhat wider confidence intervals and due to the imprecise nature of ART adherence measurement, these data are inconclusive as to whether there is an effect of alcohol on viral suppression beyond that attributed to poorer adherence.

An analysis of participants in two cohorts of a total of 595 HIV-infected persons with a history of alcohol problems examined CD4 cell counts and HIV viral loads at 6-month intervals for up to 7 years [5••]. Upon regression analysis, among subjects not on ART, heavy alcohol consumption was associated with a lower CD4 cell count, on average a difference of 49 cells/mm^3. There was no association between heavy alcohol consumption and CD4 cell count among those on ART, in analyses that adjusted for baseline CD4, adherence to ART, homelessness, depressive symptoms, and several other variables. Heavy alcohol use was not associated with HIV viral load in those on ART and those not on ART. All analyses among those on ART adjusted for 3-day self-reported adherence, suggesting that there is no detectable alcohol effect beyond the effect on adherence among those on ART. However, the CD4 cell count difference suggests that there might be an effect of heavy alcohol consumption on HIV progression among those not on ART.

A recent publication by Baum et al. [47••] examined the association between alcohol consumption and HIV outcomes in a cohort of active alcohol or illicit drug users. In this study, frequent alcohol consumption (defined as $\geq$ 2 drinks/day on average) compared to moderate alcohol use and abstention was associated with a decline of CD4 cell count to less than 200 cells/mm^3, among those who had a baseline CD4 cell count of greater than 200 cells/mm^3. The model in this study controlled for baseline CD4 cell count, HIV viral load, ART status, years since tested HIV positive, age, and gender. A similar model that examined the same factors but was restricted to those not on ART showed a stronger association. The effect size was also larger when the predictor variable was the combination of frequent alcohol use and crack cocaine use; however, the independent effects of alcohol and crack cocaine use were not shown. In addition, frequent alcohol consumption was associated with increased HIV viral load in a multivariate model controlling for the same variables as above except viral load. However, when stratifying by ART the association was significant only among those on ART, and the authors suggested that the association was mediated by adherence. These analyses are in contrast to a recent analysis of the same cohort, which found that crack cocaine use but not alcohol use, coded only as current yes versus no, was associated with HIV progression [33]. This illustrates the importance of using a more detailed alcohol consumption history to ascertain the relationship of alcohol use and HIV disease progression.

Two studies in women in the post-HAART era failed to find an association between alcohol consumption and HIV outcomes. A recent study of 516 women in the HIV Epidemiologic Research Study (HERS) cohort examined the effects of alcohol consumption (ie, none, moderate, and heavy) on both depressive symptoms and CD4 cell count [48••]. The analysis showed significant associations between both moderate and heavy alcohol consumption and depressive symptoms and between depressive symptoms and CD4 cell count. The direct association between alcohol consumption and CD4 cell count was not statistically significant. The indirect effects of alcohol consumption on CD4 cell count via depression were not reported; therefore, we cannot comment on effect of alcohol on CD4 cell count via the effect on depression. In addition, a large study of 1686 HIV-positive women in the Women's Interagency HIV Study (WIHS) found that there was no positive association between heavy alcohol consumption and time to newly acquired AIDS-defining illnesses or AIDS-related death, in repeated measures models that adjusted for crack use, ART use and adherence, CD4 cell count at baseline, HIV viral load at baseline, year of HIV diagnosis, and demographic variables [32•]. This study found a strong association of persistent crack use and AIDS-related mortality and both persistent and intermittent crack use and newly acquired AIDS illnesses. This result is consistent with other studies that included crack use in multivariate models [33, 34].

Lastly, a study conducted multivariate modeling of the effect of drink types on HIV viral load suppression, CD4 cell count, and thymus volume in 165 patients after 24 weeks of ART [49••]. In models that controlled for demographics, baseline Centers for Disease Control and Prevention HIV stage, and adherence, heavy alcohol consumption was not associated with the outcomes while consuming predominantly liquor compared to beer or wine was associated with lack of HIV viral suppression,

decreased thymus size, and change in CD4 cell count. This study highlights a potential future area of interest, that is, impact of alcohol beverage type.

In summary, we identified six studies in the post-HAART era, and three demonstrated an association between heavy alcohol use and at least one measure of HIV disease progression [5••, 24••, 47••].

Discussion

Overall, we found that there is strong biological plausibility that heavy alcohol consumption might hasten HIV disease progression. We touched on several biological mechanisms by which this might occur, including direct immunological effects of alcohol, interactions and competitions with drugs, nutritional intake, and metabolic deficiencies. In addition, we reviewed behavioral factors by which alcohol consumption could affect HIV progression, such as reduced/poor retention in HIV care, adherence, nutrition, and mental health. Given this strong evidence, one would expect that simple bivariate analyses would show strong associations between heavy alcohol consumption and HIV outcomes, and these associations would be attenuated when causal pathway variables are included in the model. However, this is not the case in most of the analyses reviewed above. Instead, pre-HAART studies showed no associations between heavy alcohol consumption and HIV outcomes, while some but not all of the later studies did find such an association. There are several possible explanations for these findings.

First, several of the early studies measured alcohol consumption at baseline, yet alcohol consumption changes over the course of HIV infection [6]. Such a misclassification of the exposure could cause an association to be obscured. However, for the studies that included time-dependent measures of heavy alcohol consumption, the feedback loop between declining health and subsequent declining alcohol consumption might have counteracted any deleterious effects of alcohol on disease progression. While some studies attempted to address that issue, by measuring changes in alcohol consumption or only measuring alcohol consumption in the first several years after diagnosis, current statistical methods such as marginal structural models may be more powerful in detecting associations [50]. This feedback loop may not have been an issue in the later studies, because ART is usually started before patients develop any outward signs of disease progression. Another issue is that some of the earlier studies focused on any or current alcohol consumption, rather than heavy alcohol consumption, which may explain the lack of associations in these studies if heavy alcohol consumption is needed to accelerate HIV disease progression.

Another possible issue is that the risk profile of the populations studied changed over time, with a shift from predominantly men who had sex with men, to inner-city clinic patients and poly-substance users. The latter groups of patients may have engaged in heavier levels of alcohol consumption or other illicit drug use which was associated with HIV disease progression.

Another possible explanation for the lack of association in the early studies is that the outcome measures shifted over time from AIDS-defining illnesses to biological markers of immunological decline (CD4) or viral replication. If alcohol has a direct effect on immune function, then it is more likely that there will be a significant association when CD4 cell count is used as the outcome variable.

Lastly, it is important to consider the possibility that publication bias became a more pervasive issue as the AIDS epidemic wore on. It seems quite conceivable that researchers evaluated associations between alcohol consumption and HIV disease outcomes in their cohorts, but did not pursue these analyses to the stage of publication if the findings were not statistically significant. Because this question is still unresolved, we suggest an analysis of existing cohort study datasets, taking into account the measurement and analysis issues raised above.

Given the ubiquitous nature of alcohol use among the people of the world who are infected with HIV, quantifying the impact of alcohol consumption on HIV disease progression has major implications on the AIDS epidemic if even only a modest effect is found. Hence, in addition to taking optimal advantage of existing data to further illuminate the relationship of HIV disease and alcohol use, identifying a cohort in which ART has not been initiated and in which alcohol is heavily consumed and can be measured would provide very valuable empirical data. Such a cohort could provide key insights into this issue, particularly if the data are collected in a manner that learns from past studies' limitations.

Conclusions

The link between alcohol use and HIV disease progression is clearly complicated to disentangle, and the more recent empirical evidence is suggestive but not strong. Although alcohol-related behavior appears to impact HIV disease progression through ART adherence, biological mechanisms are also likely to be implicated. Future studies should continue to investigate this important topic in order to provide clearer evidence, ultimately with the goals of utilizing the most valid measurement and statistical techniques and furthering our understanding by carefully controlling for confounding and meanwhile examining mechanisms of action. These studies are crucial so that

the true impact and cost-effectiveness of interventions designed to slow or prevent alcohol-associated HIV disease progression can be determined.

Disclosure No potential conflicts of interest relevant to this article were reported.

References

Papers of particular interest, published recently, have been highlighted as:
- • Of importance
- •• Of major importance

1. Samet JH, Horton NJ, Meli S, et al.: Alcohol consumption and antiretroviral adherence among HIV-infected persons with alcohol problems. Alcohol Clin Exp Res 2004, 28:572–577.
2. Conigliaro J, Gordon AJ, McGinnis KA, et al.: How harmful is hazardous alcohol use and abuse in HIV infection: do health care providers know who is at risk? J Acquir Immune Defic Syndr 2003, 33:521–525.
3. Galvan FH, Bing EG, Fleishman JA, et al.: 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, 63:179–186.
4. National Institute on Alcohol Abuse and Alcoholism: Helping patients who drink too much: a clinician's guide. Updated 2005 edition. Bethesda, MD: National Institutes of Health; 2007.
5. •• Samet JH, Cheng DM, Libman H, et al.: Alcohol consumption and HIV disease progression. J Acquir Immune Defic Syndr 2007, 46:194–199. *This longitudinal cohort study demonstrated that among HIV-infected individuals not on ART, heavy alcohol use is associated with HIV disease progression as measured by CD4 cell count, but not HIV viral load. No association was found among those on ART.*
6. Bertholet N, Cheng DM, Samet JH, et al.: Alcohol consumption patterns in HIV-infected adults with alcohol problems. Drug Alcohol Depend 2010, In press.
7. Szabo G: Consequences of alcohol consumption on host defense. Alcohol Alcohol 1999, 34:830–841.
8. Balagopal A, Philp FH, Astemborski J, et al.: Human immunodeficiency virus-related microbial translocation and progression of hepatitis C. Gastroenterology 2008, 135:226–233.
9. Fawzi W, Msamanga G, Spiegelman D, Hunter DJ: Studies of vitamins and minerals and HIV transmission and disease progression. J Nutr 2005, 135:938–944.
10. Mehta S, Fawzi W: Effects of vitamins, including vitamin A, on HIV/AIDS patients. Vitam Horm 2007, 75:355–383.
11. Lanzillotti JS, Tang AM: Micronutrients and HIV disease: a review pre- and post-HAART. Nutr Clin Care 2005, 8:16–23.
12. Drain PK, Kupka R, Mugusi F, Fawzi WW: Micronutrients in HIV-positive persons receiving highly active antiretroviral therapy. Am J Clin Nutr 2007, 85:333–345.
13. Rousseau MC, Molines C, Moreau J, Delmont J: Influence of highly active antiretroviral therapy on micronutrient profiles in HIV-infected patients. Ann Nutr Metab 2000, 44:212–216.
14. Tang AM, Forrester J, Spiegelman D, et al.: Weight loss and survival in HIV-positive patients in the era of highly active antiretroviral therapy. J Acquir Immune Defic Syndr 2002, 31:230–236.
15. Tang AM, Jacobson DL, Spiegelman D, et al.: Increasing risk of 5% or greater unintentional weight loss in a cohort of HIV-infected patients, 1995 to 2003. J Acquir Immune Defic Syndr 2005, 40:70–76.
16. Quach LA, Wanke CA, Schmid CH, et al.: Drug use and other risk factors related to lower body mass index among HIV-infected individuals. Drug Alcohol Depend 2008, 95:30–36.
17. Scrimshaw NS, SanGiovanni JP: Synergism of nutrition, infection, and immunity: an overview. Am J Clin Nutr 1997, 66:464S–477S.
18. Watzl B, Watson RR: Role of alcohol abuse in nutritional immunosuppression. J Nutr 1992, 122(3 Suppl):733–737.
19. Molina PE, McNurlan M, Rathmacher J, et al.: Chronic alcohol accentuates nutritional, metabolic, and immune alterations during asymptomatic simian immunodeficiency virus infection. Alcohol Clin Exp Res 2006, 30:2065–2078.
20. Lieber CS, DeCarli LM: Hepatic microsomal ethanol-oxidizing system. In vitro characteristics and adaptive properties in vivo. J Biol Chem 1970, 245:2505–2512.
21. Lieber CS: Medical and nutritional complications of alcoholism: mechanisms and management. New York: Plenum Press; 1992.
22. • Conen A, Fehr J, Glass TR, et al.: Self-reported alcohol consumption and its association with adherence and outcome of antiretroviral therapy in the Swiss HIV Cohort Study. Antivir Ther 2009, 14:349–357. *This large (n= 6323) cross-sectional study demonstrated the relationship between heavy drinking and poorer ART adherence.*
23. Martinez P, Andia I, Emenyonu N, et al.: Alcohol use, depressive symptoms and the receipt of antiretroviral therapy in southwest Uganda. AIDS Behav 2008, 12:605–612.
24. •• Chander G, Lau B, Moore RD: Hazardous alcohol use: a risk factor for non-adherence and lack of suppression in HIV infection. J Acquir Immune Defic Syndr 2006, 43:411–417. *This prospective study found that heavy alcohol use alone and in combination with other drug use was independently associated with lower odds of ART use, adherence, and viral suppression.*
25. Cunningham WE, Sohler NL, Tobias C, et al.: Health services utilization for people with HIV infection: comparison of a population targeted for outreach with the U.S. population in care. Med Care 2006, 44:1038–1047.
26. Sabin CA, Smith CJ, Gumley H, et al.: Late presenters in the era of highly active antiretroviral therapy: uptake of and responses to antiretroviral therapy. AIDS 2004, 18:2145–2151.
27. • Hendershot CS, Stoner SA, Pantalone DW, Simoni JM: Alcohol use and antiretroviral adherence: review and meta-analysis. J Acquir Immune Defic Syndr 2009, 52:180–202. *The authors identified 40 studies to include in a meta-analysis to evaluate the relationship between alcohol and ART adherence. The results reveal a significant and consistent association between alcohol and worse ART adherence.*
28. Braithwaite RS, McGinnis KA, Conigliaro J, et al.: A temporal and dose-response association between alcohol consumption and medication adherence among veterans in care. Alcohol Clin Exp Res 2005, 29:1190–1197.
29. Sullivan LE, Saitz R, Cheng DM, et al.: The impact of alcohol use on depressive symptoms in human immunodeficiency virus-infected patients. Addiction 2008, 103:1461–1467.
30. Leserman J: Role of depression, stress, and trauma in HIV disease progression. Psychosom Med 2008, 70:539–545.

31. Gore-Felton C, Koopman C: Behavioral mediation of the relationship between psychosocial factors and HIV disease progression. Psychosom Med 2008, 70:569–574.
32. • Cook JA, Burke-Miller JK, Cohen MH, et al.: Crack cocaine, disease progression, and mortality in a multicenter cohort of HIV-1 positive women. AIDS 2008, 22:1355–1363. *This longitudinal assessment of HIV-infected US women showed that crack cocaine but not heavy alcohol use was an independent predictor of AIDS-related morbidity and mortality and HIV disease progression as measured by CD4 cell count and HIV viral load*.
33. Baum MK, Rafie C, Lai S, et al.: Crack-cocaine use accelerates HIV disease progression in a cohort of HIV-positive drug users. J Acquir Immune Defic Syndr 2009, 50:93–99.
34. Kumar R, Perez-Casanova AE, Tirado G, et al.: Increased viral replication in simian immunodeficiency virus/simian-HIV-infected macaques with self-administering model of chronic alcohol consumption. J Acquir Immune Defic Syndr 2005, 39:386–390.
35. Marcondes MC, Watry D, Zandonatti M, et al.: Chronic alcohol consumption generates a vulnerable immune environment during early SIV infection in rhesus macaques. Alcohol Clin Exp Res 2008, 32:1583–1592.
36. Bagby GJ, Zhang P, Purcell JE, et al.: Chronic binge ethanol consumption accelerates progression of simian immunodeficiency virus disease. Alcohol Clin Exp Res 2006, 30:1781–1790.
37. Poonia B, Nelson S, Bagby GJ, et al.: Chronic alcohol consumption results in higher simian immunodeficiency virus replication in mucosally inoculated rhesus macaques. AIDS Res Hum Retroviruses 2006, 22:589–594.
38. Kaslow RA, Blackwelder WC, Ostrow DG, et al.: No evidence for a role of alcohol or other psychoactive drugs in accelerating immunodeficiency in HIV-1-positive individuals. A report from the Multicenter AIDS Cohort Study. JAMA 1989, 261:3424–3429.
39. Penkower L, Dew MA, Kingsley L, et al.: Alcohol consumption as a cofactor in the progression of HIV infection and AIDS. Alcohol 1995, 12:547–552.
40. van Griensven GJ, de Vroome EM, de Wolf F, et al.: Risk factors for progression of human immunodeficiency virus (HIV) infection among seroconverted and seropositive homosexual men. Am J Epidemiol 1990, 132:203–210.
41. Eskild A, Petersen G: Cigarette smoking and drinking of alcohol are not associated with rapid progression to acquired immunodeficiency syndrome among homosexual men in Norway. Scand J Soc Med 1994, 22:209–212.
42. Tang AM, Graham NM, Chandra RK, Saah AJ: Low serum vitamin B-12 concentrations are associated with faster human immunodeficiency virus type 1 (HIV-1) disease progression. J Nutr 1997, 127:345–351.
43. Veugelers PJ, Page KA, Tindall B, et al.: Determinants of HIV disease progression among homosexual men registered in the Tricontinental Seroconverter Study. Am J Epidemiol 1994, 140:747–758.
44. Webber MP, Schoenbaum EE, Gourevitch MN, et al.: A prospective study of HIV disease progression in female and male drug users. AIDS 1999, 13:257–262.
45. Crum RM, Galai N, Cohn S, et al.: Alcohol use and T-lymphocyte subsets among injection drug users with HIV-1 infection: a prospective analysis. Alcohol Clin Exp Res 1996, 20:364–371.
46. Chandiwana SK, Sebit MB, Latif AS, et al.: Alcohol consumption in HIV-I infected persons: a study of immunological markers, Harare, Zimbabwe. Cent Afr J Med 1999, 45:303–308.
47. •• Baum MK, Rafie C, Lai S, et al.: Alcohol use accelerates HIV disease progression. AIDS Res Hum Retroviruses 2010, 26:511–518. *This recent work was consistent with the study by Samet et al. [5••] that found that alcohol use is associated with a decrease in CD4 cell count; however, this study further found this association overall and among those not on ART. In addition, the authors found that alcohol use was associated with higher HIV viral load, only among those on ART*.
48. •• Ghebremichael M, Paintsil E, Ickovics JR, et al.: Longitudinal association of alcohol use with HIV disease progression and psychological health of women with HIV. AIDS Care 2009, 21:834–841. *This longitudinal study of 871 HIV-infected US women found no significant association between alcohol use and CD4 cell count, regardless of ART use. This work did find an association between alcohol use and depression and between depression and CD4 cell count*.
49. •• Miguez-Burbano MJ, Lewis JE, Fishman J, et al.: The influence of different types of alcoholic beverages on disrupting highly active antiretroviral treatment (HAART) outcome. Alcohol Alcohol 2009, 44:366–371. *This prospective study examined the effect of beverage type on several HIV outcomes among persons on ART. They found that those consuming primarily liquor as compared to beer or wine had poorer outcomes, even after adjusting for number of drinks and ART adherence*.
50. Bryan J, Yu Z, Van Der Laan MJ: Analysis of longitudinal marginal structural models. Biostatistics 2004, 5:361–380.

Longitudinal Patterns in Survival, Comorbidity, Healthcare Utilization and Quality of Care among Older Women Following Breast Cancer Diagnosis

Amresh D. Hanchate, PhD[1,2], Kerri M. Clough-Gorr, DSc, MPH[3,4,5], Arlene S. Ash, PhD[2,6], Soe Soe Thwin, PhD[3,7], and Rebecca A. Silliman, MD, PhD[3,4]

[1]Center for Organization, Leadership and Management Research (COLMR), VA Boston Healthcare System, Boston, MA, USA; [2]Health Care Research Unit, Section of General Internal Medicine, Boston University School of Medicine, Boston, MA, USA; [3]Section of Geriatrics, Boston University School of Medicine, Boston, MA, USA; [4]Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA; [5]National Institute for Cancer Epidemiology and Registration (NICER) Institute of Social and Preventative Medicine (ISPM), University of Zurich, Zurich, Switzerland; [6]Department of Quantitative Health Sciences, University of Massachusetts Medical School, Worcester, MA, USA; [7]Massachusetts Veterans Epidemiology Research and Information Center (MAVERIC), VA Healthcare System, Boston, MA, USA.

OBJECTIVES: To compare longitudinal patterns of health care utilization and quality of care for other health conditions between breast cancer-surviving older women and a matched cohort without breast cancer.

DESIGN: Prospective five-year longitudinal comparison of cases and matched controls.

SUBJECTS: Newly identified breast cancer patients recruited during 1997–1999 from four geographic regions (Los Angeles, CA; Minnesota; North Carolina; and Rhode Island; N=422) were matched by age, race, baseline comorbidity and zip code location with up to four non-breast-cancer controls (N=1,656).

OUTCOMES: Survival; numbers of hospitalized days and physician visits; total inpatient and outpatient Medicare payments; guideline monitoring for patients with cardiovascular disease and diabetes, and bone density testing and colorectal cancer screening.

RESULTS: Five-year survival was similar for cases and controls (80% and 82%, respectively; p=0.18). In the first follow-up year, comorbidity burden and health care utilization were higher for cases (p<0.01), with most differences diminishing over time. However, the number of physician visits was higher for cases (p<0.01) in every year, driven partly by more cancer and surgical specialist visits. Cases and controls adhered similarly to recommended bone density testing, and monitoring of cardiovascular disease and diabetes; adherence to recommended colorectal cancer screening was better among cases.

CONCLUSION: Breast cancer survivors' health care utilization and disease burden return to pre-diagnosis levels after one year, yet their greater use of outpatient care persists at least five years. Quality of care for other chronic health problems is similar for cases and controls.

Received September 30, 2009
Revised March 31, 2010
Accepted May 12, 2010
Published online June 8, 2010

KEY WORDS: survival; case-control; inpatient care; outpatient care; costs; preventive care.

J Gen Intern Med 25(10):1045–50
DOI: 10.1007/s11606-010-1407-9

Demographics and therapeutic progress in the United States are increasing the number of older cancer survivors. About 182,500 women were diagnosed with breast cancer in 2008, almost half occurring in women aged 65 or older. At five years, 89% remain alive;[1] of over 2.6 million breast cancer survivors in the United States, more than half are ≥65 years old.[2] Both the prevalence and absolute numbers of breast cancer survivors will grow, because aging is the most important risk factor for breast cancer;[3] gains in life expectancy and advances in detection and treatment will place more women at risk for breast cancer, and breast cancer survivorship.

Comorbidities also increase with age, for those with cancer and others.[4–7] Across most cancer types, older cancer patients report significantly more comorbidity and poorer physical health than non-cancer patients of the same age.[6–8] Since hypertension, heart conditions, arthritis, and diabetes are common in older breast cancer survivors,[7] the Institute of Medicine has emphasized quality follow-up care for this cohort.[9,10]

Although follow-up care is essential in cancer survivorship,[11–15] studies comparing the quality of care for cancer survivors to those without cancer are conflicting. A longitudinal study of patients with diabetes showed similar quality of diabetes care for patients with and without cancer.[16] Other studies have found that older cancer survivors receive similar, or less follow-up care than controls, or that the type of primary and cancer-related care varies by provider type.[9,11–14] Finally,

a SEER-Medicare study of older breast cancer survivors found them receiving more preventative services than non-cancer patients.[17] The inconsistent results may be due to different study populations (e.g. source, age, and type of cancer) and/or the kinds of follow-up care examined.

To answer whether overall healthcare utilization and quality of follow-up care for other conditions differs for older women with and without breast cancer, we conducted a multi-site study of women ≥65-years of age with breast cancer and age-, morbidity-, and geography-matched controls. We compared differences in survival and quality of follow-up care over five years.

METHODS

Study Design

In this prospective, longitudinal study, we compared survival, comorbidity, healthcare utilization, and quality of care between breast cancer patients and a matched control cohort using Medicare data from the Centers for Medicare and Medicaid Services (CMS). We examined annual comorbidity and healthcare utilization for five years following the date of breast cancer diagnosis. Quality of care was defined as guideline-consistent colorectal cancer screening, bone density testing, and monitoring of cardiovascular disease and diabetes.

Study Participants

Recruitment procedures for the breast cancer cohort are detailed elsewhere.[18] Briefly, we identified newly diagnosed breast cancer patients by reviewing pathology reports at hospitals or tumor registries in four geographic regions (Los Angeles, California; Minnesota; North Carolina; and Rhode Island). Each setting received Institutional Review Board (IRB) approval. Eligibility required: stage II–IIIA disease or stage I disease with a tumor diameter ≥1 cm; age at diagnosis ≥65 years; no prior history of primary breast cancer; no simultaneously diagnosed or treated second primary tumor at another site; English speaking; and, competent for interview.

Eligible participants were mailed an enrollment package and called by a research staff member. Between 1997 and 1999, 921 women agreed to participate by returning an IRB-approved signed consent form, and were enrolled. Tumor information was collected by medical record review, with date of definitive biopsy treated as the enrollment date for each case. Based on additional information, 56 respondents did not meet inclusion criteria, leaving 865 eligible cases.

Because fee-for-service CMS records were the only source of utilization information for both cases and controls, some otherwise eligible cases could not be included in this comparison study. Such records were only available for Medicare enrollees with both inpatient (Part A) and outpatient (Part B) entitlement and fee-for-service (FFS) coverage. We excluded 301 cases without such coverage for the entire period beginning one year before study enrollment and for five years following. Additional case exclusions were due to problematic Medicare identifiers (n=136); no matched controls (n=5); and no confirmation of breast cancer diagnosis in the administrative data (n=1).

For the remaining 422 cases, we selected controls in two stages. First, using the full (100%) Medicare data, we identified all potential controls (n=21,241) that exactly matched at least one case by age, race (White, Black or Other) and 5-digit zip code location. For 5.8% of cases we did not find exact matches and therefore relaxed the matching criteria for age (±/−2-years) and zip code (within 20-mile radius). We excluded women with a history of breast cancer.

From among the 21,241 potential controls for the 422 cases, we sought four controls for each case—specifically the four non-cases who best matched the comorbidity burden of the case in the year prior to her cancer diagnosis. Selected controls were assigned the same "enrollment date" as its matched case.

"Comorbidity burden" was measured by a prospective diagnostic cost group (DCG) score, calculated from age, sex and all ICD-9-CM diagnosis codes recorded—excluding breast cancer diagnosis codes—for inpatient and outpatient encounters during the year preceding "enrollment."[19] The diagnostic cost group (DCG) risk adjustment system is a validated risk adjustment system.[20,21] Originally developed for setting prospective Medicare payments, DCG models are now used to risk adjust various outcomes in a range of populations. The DCG score used here indicates the expected future cost of utilization, normalized to average 1.0 in the Medicare over-age-65 population. All but 13 of the 422 cases had at least five potential controls, from which the four potential controls with comorbidity burden (DCG score) nearest to that of the matched case were selected. Of the rest, one had four potential controls; two had three; and ten had one. These were all included in the final study population of 422 cases and 1,656 controls.

Data Sources and Variables

For the study population, we obtained Medicare eligibility files for 1996 to 2004 and Medicare claims files (MEDPAR, carrier and outpatient) for 1996 to 2003.

We defined outcome measures using the day after the study enrollment date (which differs by case), as the beginning of follow-up. We identified dates of death through 12/31/2004 from CMS denominator files. We measured comorbidity burden using the DCG score, as above, for each of five 365-day periods subsequent to the study enrollment date. To measure total *non-breast cancer* illness burden, we dropped breast cancer codes (ICD-9-CM 174.0-174.9, 198.81) before calculating the score. In each year, utilization was measured as follows: overall utilization as total Medicare payments; inpatient, as both 1) number of hospital days and 2) Medicare payments charged for inpatient care: outpatient, as payments for outpatient care (including physician visits, imaging, laboratory tests and procedures). We also examined payments by type of care,[22] and a narrow outpatient measure: number of physician visits. Using provider specialty codes we categorized visits by specialty—cancer, cardiopulmonary, mental health, surgery, generalist and other. Since outliers among individual expenditures would unduly influence overall statistics, annual inpatient payment measures above $50,000 were reset to $50,000; similarly, outpatient payment measures were top-coded at $25,000. This "top-coding" affected at most eight observations for any measure.

Quality of care was measured by adherence to a) guideline-consistent colorectal cancer screening[23] and b) bone density testing for all subjects[24], and c) recommended monitoring for those with cardiovascular disease (CVD) or with diabetes (DM),

identified from ICD-9-CM diagnosis codes prior to the enrollment date.

Analyses

We used baseline interview data to compare breast cancer patients included in this study (N=422) with the cases (N=443) that could not be included. The groups were similar with respect to age, race/ethnicity, and a range of comorbidity, tumor and treatment characteristics.

Our key comparisons were of the 422 cases and 1,656 controls, using t-tests, chi-square tests, and log-rank tests (for survival data) as appropriate. We used Kaplan-Meier survival analyses, with log-rank tests for differences by group, and proportional hazards regression models to examine group differences while controlling for differences in baseline comorbidity, age, race and geography between cases and controls.

For annual utilization measures, we report t-test comparisons between cohorts. Analogous adjusted comparisons were tested using linear multivariate regression. Given that the cohorts were fairly well matched on important predictors, regression-based findings were very like those based on t-tests. Thus, we only report the latter. Quality of care measures were compared using chi-square tests.

RESULTS

Table 1 characterizes the demographics and baseline comorbidity burden of the cases and the matched controls. In both groups, over half were aged 65 to 74; about 4% were black; over 70% of both cohorts had less comorbidity at baseline than the Medicare average (1.0), mainly due to younger ages. While not statistically significant, more cases (6.9%) than controls (5.0%) were in the highest comorbidity burden category (DCG >3.0).

Most subjects (74% of cases and 76% of controls) were alive as of 12/31/2004, the study end date (p=0.16). Five-year survival was 80% and 82% for cases and controls, respectively (p=0.18). Annual mortality (Fig. 1) was also similar by either log-rank test or Cox proportional hazards regression (both $p \geq 0.15$).

Table 1. Characteristics of the Study Population

	Cases (n=422)	Controls (n=1,656)
Age at study inception[a] (%)		
65–74	54.5	54.3
75–84	37.2	37.4
85+	8.3	8.3
% Black	4.3	4.4
Baseline comorbidity burden score[b] (DCG score) %		
<1.0	73.7	75.7
1.0 to 2.0	19.4	19.3
>3.0	6.9	5.0
Year of study inception[a]		
1997	19.7	20.0
1998	54.5	54.2
1999	25.8	25.8

Similarity between the two cohorts was not rejected for any of the characteristics (at 5% significance level)
[a]*Based on date of diagnosis for cases; for controls, the date of diagnosis for the matched case was used*
[b]*Based on the 1-year prior to study inception*

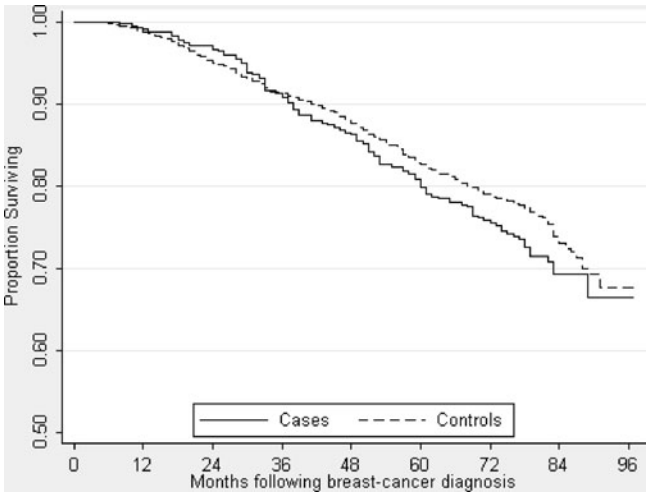


Figure 1. Survivorship of breast cancer cases and controls.

Table 2 compares healthcare utilization and comorbidity burden annually for five years following enrollment. During the first year following breast cancer diagnosis, Medicare payments for cases ($11,193) were significantly higher than for controls ($3,159; p<0.01). The magnitude of this difference decreased in subsequent years, yet Medicare payments remained significantly higher for cases in three of the four years. Non-breast cancer morbidity was higher for cases than controls in all years.

Table 3 characterizes utilization differences separately for inpatient and outpatient care. In the year following breast cancer diagnosis, average Medicare payments for cases were $3,935 for inpatient care and $7,259 for outpatient care. Both were notably higher than for controls ($1,710 and $1,450, respectively, p<0.01). This difference is reflected in the quantity measures: first-year hospital days averaged 3.9 for cases and 1.7 for controls (p<0.01), and average number of outpatient physician visits for cases (16.6) was twice that for controls (8.1; p<0.01). However, trends differed markedly for outpatient and inpatient care. The number of outpatient physician visits remained significantly higher for cases in each of the four subsequent years, although the year-five difference was smallest (cases=11.0; controls=8.8, p<0.01). In all five years, there were similar visit numbers for generalists and most specialists, but higher visit numbers for cancer and surgery specialists. In terms of Medicare payments, this difference was significant (in all but the fifth year). For outpatient payments by type of care—physician visits, imaging, laboratory tests and procedures—differences were uniform across all types. In contrast, differences in inpatient care, measured either as Medicare payments or hospital days, were not significantly different in three of the four years.

We also examined quality of care for all patients and for those with CVD or DM at study entry, which were similarly prevalent among cases and controls (Table 4). Among those with CVD, missing at least one biennial lipid test prior to the end of 2003 (for survivors) or the date of death (for non-survivors), was more likely than not for both cases and controls. Diabetes was also similarly prevalent (19% vs. 17%, for cases and controls, respectively), as were lipid and AIC testing, and eye examinations. Among all cases and controls,

Table 2. Total Medicare Payments and Comorbidity Burden Following Breast Cancer Diagnosis

Period after diagnosis	Total Medicare payments ('000 $)			Comorbidity burden (DCG) score		
	Mean (Std Dev)		P-value	Mean (Std Dev)		P-value
	Cases	Controls		Cases	Controls	
1st Year	11.2 (8.3)	3.2 (7.2)	<0.01	0.90 (0.84)	0.80 (0.68)	0.01
2nd Year	3.9 (7.0)	3.3 (7.1)	0.10	1.04 (1.01)	0.96 (0.87)	0.12
3rd Year	5.1 (8.2)	3.5 (7.0)	<0.01	1.11 (1.00)	0.98 (0.87)	0.01
4th Year	5.4 (9.1)	4.0 (7.7)	<0.01	1.15 (1.03)	1.06 (1.01)	0.12
5th Year	3.6 (5.6)	3.6 (7.9)	0.96	0.95 (0.76)	0.95 (0.82)	0.90

a Comorbidity burden, excluding breast cancer, is measured by diagnosis-based DCG score. Score=1 denotes mean Medicare enrollee burden
b P-value is from a t-test for equal means in same-year measure between cases and controls

bone density testing was also equally prevalent; however, non-adherence to colorectal cancer screening among cases (33%) was lower than that for controls (47%; p-value < 0.001).

DISCUSSION

This prospective study contrasted health outcomes and healthcare utilization for a cohort of breast cancer patients over five years following their cancer diagnosis with those of controls matched by age, geographic location and total comorbidity burden. Non-breast cancer comorbidity at baseline and five-year survival were similar in the two cohorts at baseline. Although similar in all other years, non-breast cancer comorbidity was significantly higher for cases in the first year following breast cancer diagnosis. Among survivors in each cohort, hospital utilization exhibited the same pattern, with more days of hospitalization and Medicare payments in the first follow-up year, but similar amounts subsequently. In contrast, both numbers of physician visits and total Medicare payments for outpatient services (as measured by Medicare payments) were notably higher for cases—by over 25%—during most of the five follow-up years. Quality of care, as measured by guideline-adherent monitoring for those with baseline CVD and DM, and bone density testing for all patients, was also similar in the groups. However, colorectal cancer screening guideline adherence was better for cases.

Our estimates of utilization following breast cancer diagnosis are like those of Yabroff et al. in which Medicare payments in the first year following breast cancer diagnosis were $11,728 (1999–2003), and higher than for a cohort matched by age and geographic location, but not comorbidity.[25] While the corresponding estimate from our study was $8,034, these figures are not directly comparable due to differences in control selection, study time-period, and geography. However, we estimated inpatient costs for our cases to be 27% of total Medicare payments for the cases, which is similar to the Yabroff et al. estimate of 25%. Note that Medicare payments serve as a good proxy for inpatient and outpatient health care services as Medicare is the primary payer, accounting for over 86% of such expenditures for the elderly.[26]

Some studies have found greater comorbidity among cancer patients than among age-similar non-cancer patients.[5–7] Therefore, when breast cancer patients are compared with non-cancer patients matched only on age and location,[25,27,28] the cancer patients could have higher healthcare utilization for this reason alone. In this study we matched by baseline comorbidity in addition to age, race and geography; this enabled us to ask if the breast cancer diagnosis and treatment itself have longer term effects on healthcare utilization? While inpatient care utilization was similar between the two cohorts, numbers of physician visits and Medicare payments for outpatient care were consistently greater among breast cancer survivors—by over 100% in the first year and at least 25% in all four subsequent years. Breast cancer survivors visited

Table 3. Inpatient and Outpatient Healthcare Use Following Breast Cancer Diagnosis

Period after diagnosis	Mean (Std Dev)		P-value	Mean (Std Dev)		P-value
	Cases	Controls		Cases	Controls	
Inpatient care						
	Inpatient Medicare payments ('000 $)			# Hospitalization days		
1st Year	3.9 (6.2)	1.7 (5.5)	<0.01	3.9 (6.8)	1.7 (5.2)	<0.01
2nd Year	1.7 (5.2)	1.7 (5.5)	0.99	1.8 (5.4)	1.6 (5.1)	0.60
3rd Year	2.5 (6.4)	1.7 (5.2)	0.01	2.2 (6.1)	1.7 (5.3)	0.09
4th Year	2.4 (6.6)	1.9 (5.7)	0.17	2.2 (6.4)	1.8 (5.7)	0.26
5th Year	1.4 (4.1)	1.6 (4.9)	0.65	1.2 (3.7)	1.4 (4.4)	0.59
Outpatient care						
	All outpatient Medicare payments ('000 $)			# Physician Visits		
1st Year	7.3 (4.8)	1.5 (2.2)	<0.01	16.6 (11.1)	8.1 (8.5)	<0.01
2nd Year	2.2 (2.7)	1.5 (2.3)	<0.01	12.5 (10.1)	8.6 (9.1)	<0.01
3rd Year	2.6 (3.1)	1.8 (2.6)	<0.01	12.3 (9.9)	8.9 (9.7)	<0.01
4th Year	3.0 (4.0)	2.1 (3.0)	<0.01	13.0 (10.5)	9.6 (9.6)	<0.01
5th Year	2.2 (2.7)	2.1 (1.8)	0.57	11.0 (9.5)	8.8 (9.2)	<0.01

a P-value is from a t-test of equal mean in same-year measures between cases and controls

Table 4. Comparison of Indicators of Inadequate Quality of Chronic Disease Care

	Survivors	Controls	*P*-value
Cardiovascular disease (CVD)			
Baseline prevalence, %	49	49	0.97
CVD patients who missed biennial lipid test, %	49	48	0.87
Diabetes (DM)			
Baseline prevalence, %	19	17	0.36
DM patients who missed biennial lipid test, %	39	39	0.98
DM patients who missed annual A1C test, %	66	64	0.74
DM patients who missed biennial eye exam, %	42	43	0.80
All			
% missed guideline colorectal cancer screening during follow-up	21	27	0.03
% with no bone density testing during follow-up	57	59	0.47

Data on care received following breast cancer diagnosis till Dec 31, 2003
a The follow-up period extends from the day after enrollment to 12/31/2003 or the date of death if the latter is earlier. Using the U.S. Preventive Services Taskforce guideline for colorectal cancer screening, we defined compliance for this study cohort as one of the following during the follow-up period (for subjects in 65–75 year age group): i) annual screening with high-sensitivity fecal occult blood testing, ii) one sigmoidoscopy examination and one fecal occult blood test, and iii) one colonoscopy examination[23,39]

cancer specialists more often, but numbers of visits to mental health specialists and generalists were similar. Some of this difference may be due to a "volunteer effect", in which those who agree to enroll in a study are more likely to engage in systematic care-seeking than the controls who were selected by retrospective matching; the control group contains both women who would have and those who would not have accepted study enrollment had it been offered.[29,30]

Greater outpatient care among breast cancer survivors could also reflect long-term and late complications of cancer treatment,[13,31] since comorbid conditions and their treatment may interact with cancer and its treatment, leading to worse physical health and higher mortality risk.[6,32,33] However, survival for at least five years following diagnosis was quite similar for cases and controls, as was measured comorbidity levels in most years. Hence, more physician visits and outpatient care among breast cancer survivors does not appear to be due to their being sicker. More likely is that increased outpatient care is associated with breast cancer follow-up care. Indeed, the Institute of Medicine report *From Cancer Patient to Cancer Survivor: Lost in Transition* focuses on the complex follow-up needs of cancer survivors, including preventive care, monitoring for treatment side effects (e.g., adjuvant hormonal therapy) and recurrence surveillance.[9]

That is, breast cancer patients, after being drawn into the caregiving network, are likely to remain engaged. Several studies have examined whether breast cancer modifies care-seeking for comorbid conditions among older adults.[11,16,17,34–36] We examined colorectal cancer screening and bone density testing for all patients, and monitoring for two important chronic condition subgroups—those with CVD and DM. We found no differences between cases and controls, except in the case of colorectal cancer screening. Similar or better quality of care among breast cancer survivors is consistent with our finding that they had more physician visits compared to the controls. Differing study design makes comparisons with previous reports difficult. Snyder et al. found that in each of the five years of follow-up, breast cancer survivors had less colorectal cancer screening, bone density and lipid testing than matched controls.[36] However, their controls were chosen from women who had had a mammogram during the baseline year, making it likely that their overall quality of care was also above average. In contrast, Earle et al. found that breast cancer survivors with diabetes had higher rates of lipid testing than matched controls.[17] Since comorbidity was not a matching criterion, this could have been due to differences in comorbid disease burden. Ultimately it has not been clear whether breast cancer survivors receive either more or less chronic disease care than similarly-ill women with the same morbidity burden. Keating et al., which also matched controls by comorbidity, found any-cancer survivors with diabetes receiving diabetes screening "of generally similar quality" as non-cancer diabetics.[16] This is consistent with our findings.

Previous studies have examined breast cancer survivors retrospectively;[17,37] in contrast, we have been able to examine survival prospectively. The estimated five-year survival for this breast cancer cohort (79.6%), is lower than that for National Cancer Institute's Surveillance Epidemiology and End Results (SEER) regions (89.6%; 1996–2003).[38] This could partly be due to the difference in populations represented—our subjects were from four selected geographic areas, while SEER data is obtained from areas containing over 26% of the national population. Further, because we were also interested in studying utilization, we examined only fee-for-service Medicare beneficiaries.

This study has several limitations. The study population was clustered in four geographic areas and may not generalize nationally. The breast cancer cohort had volunteered for a study; they may be healthier, or more prone to positive health behaviors, than other survivors, or a matched cohort selected from the general population.[29,30] Also, we used only Medicare administrative data, with limited clinical detail on comorbidity severity and no pharmacy data; since patient comorbidity was identified from diagnoses coded on claim forms, breast cancer survivors' greater interaction with health care providers may partly explain their higher measured comorbidity. Both cases and controls participated in Medicare FFS; in this sense they were similar, but not fully representative of Medicare enrollees, upwards of 20% of whom are enrolled in managed care plans. Also, since studying survivorship requires waiting for the data to mature, our subjects' care does not reflect recent advances in breast cancer management, such as sentinel lymph node biopsy and the use of aromatase inhibitors. We also recognize that the sub-cohorts examined for quality of care, involving those with baseline CVD and diabetes, were small. The study cohort has few minorities (4.3% black), a result of a) the population distribution of the study regions and b) older black women being at lower risk of developing breast cancer, but at higher risk of presenting with late stage disease (and therefore excluded from our study). Finally, due to lack of completeness of provider specialty field, the provider type of a sizable proportion of physician visits (about 25%) could not be determined.

In summary, this study clarifies the mixed picture related to longitudinal health care for older breast cancer survivors. We found that, beyond the first year after breast cancer diagnosis, older survivors have patterns of disease burden, inpatient care and quality of care for other health problems quite similar to

those of women without breast cancer. There was one exception—they had notably more physician visits, especially for cancer and surgical specialists, than non-breast-cancer patients in each of five years of follow-up. Reasons for this should be pursued using more detailed clinical data. For example, were the excess visits for cancer surveillance, and, if so, did they follow accepted guidelines?

Acknowledgements: *This work was supported by grants CA106979, CA/AG 70818, CA84506, and CA92395 from the National Cancer Institute.*

Conflict of Interest: *None disclosed.*

Corresponding Author: *Amresh D. Hanchate, PhD; Health Care Research Unit, Section of General Internal Medicine, Boston University School of Medicine, 801, Massachusetts Ave #2077, Boston, MA 01532, USA (e-mail: hanchate@bu.edu).*

REFERENCES

1. American Cancer Society. Breast Cancer: Facts & Figures. 2007.
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(17):1784–92.
3. Natinal Cancer Institute. Probability of Breast Cancer in American Women. 2009. http://www.cancer.gov/cancertopics/factsheet/Detection/probability-breast-cancer. Published Last Modified Date|. Accessed May 2010|.
4. **Guralnik JM.** Assessing the impact of comorbidity in the older population. Ann Epidemiol. 1996;6(5):376–80.
5. **Ogle KS, Swanson GM, Woods N, Azzouz F.** Cancer and comorbidity: redefining chronic diseases. Cancer. 2000;88(3):653–63.
6. **Smith AW, Reeve BB, Bellizzi KM, et al.** Cancer, comorbidities, and health-related quality of life of older adults. Health Care Financ Rev. 2008;29(4):41–56.
7. **Yancik R, Havlik RJ, Wesley MN, et al.** Cancer and comorbidity in older patients: a descriptive profile. Ann Epidemiol. 1996;6(5):399–412.
8. **Keating NL, Norredam M, Landrum MB, Huskamp HA, Meara E.** Physical and mental health status of older long-term cancer survivors. J Am Geriatr Soc. 2005;53(12):2145–52.
9. **Hewitt ME, Greenfield S, Stovall E,** National Cancer Policy Board (U.S.). Committee on Cancer Survivorship: Improving Care and Quality of Life. From Cancer Patient to Cancer Survivor: Lost in Transition. Washington: National Academies Press; 2006.
10. **Snyder CF, Earle CC, Herbert RJ, Neville BA, Blackford AL, Frick KD.** Preventive care for colorectal cancer survivors: a 5-year longitudinal study. J Clin Oncol. 2008;26(7):1073–9.
11. **Burstein HJ, Winer EP.** Primary care for survivors of breast cancer. N Engl J Med. 2000;343(15):1086–94.
12. **Hurria A, Hudis C.** Follow-up care of breast cancer survivors. Crit Rev Oncol Hematol. 2003;48(1):89–99.
13. **Partridge AH, Winer EP, Burstein HJ.** Follow-up care of breast cancer survivors. Semin Oncol. 2003;30(6):817–25.
14. **Earle CC.** Failing to plan is planning to fail: improving the quality of care with survivorship care plans. J Clin Oncol. 2006;24(32):5112–6.
15. Institute of Medicine. From cancer patient to cancer survivor: Lost in transition. Washington: National Academy Press; 2005.
16. **Keating NL, Zaslavsky AM, Herrinton LJ, Selby JV, Wolf RE, Ayanian JZ.** Quality of diabetes care among cancer survivors with diabetes. Med Care. 2007;45(9):869–75.
17. **Earle CC, Burstein HJ, Winer EP, Weeks JC.** Quality of non-breast cancer health maintenance among elderly breast cancer survivors. J Clin Oncol. 2003;21(8):1447–51.
18. **Silliman RA, Guadagnoli E, Rakowski W, et al.** Adjuvant tamoxifen prescription in women 65 years and older with primary breast cancer. J Clin Oncol. 2002;20(11):2680–8.
19. DxCG Inc. DxCG RiskSmart: Clinical Classifications Guide. Boston: DxCG; 2007.
20. **Ash AS, Ellis RP, Pope GC, et al.** Using diagnoses to describe populations and predict costs. Health Care Financ Rev. 2000;21(3):7–28.
21. **Ash AS, Posner MA, Speckman J, Franco S, Yacht AC, Bramwell L.** Using claims data to examine mortality trends following hospitalization for heart attack in Medicare. Health Serv Res. 2003;38(5):1253–62.
22. **Brown ML, Riley GF, Schussler N, Etzioni R.** Estimating health care costs related to cancer treatment from SEER-Medicare data. Med Care. 2002;40(8 Suppl):IV-104–17.
23. **Whitlock EP, Lin JS, Liles E, Beil TL, Fu R.** Screening for colorectal cancer: a targeted, updated systematic review for the U.S. Preventive Services Task Force. Ann Intern Med. 2008;149(9):638–58.
24. U.S. Preventive Services Task Force. Screening for Osteoporosis in Postmenopausal Women: Recommendations and Rationale. Rockville: Agency for Healthcare Research and Quality; 2002.
25. **Yabroff KR, Lamont EB, Mariotto A, et al.** Cost of care for elderly cancer patients in the United States. J Natl Cancer Inst. 2008;100(9):630–41.
26. Centers for Medicare & Medicaid Services. National Health Expenditures Fact Sheet: U.S. Department of Health & Human Services; 2009.
27. **Warren JL, Brown ML, Fay MP, Schussler N, Potosky AL, Riley GF.** Costs of treatment for elderly women with early-stage breast cancer in fee-for-service settings. J Clin Oncol. 2002;20(1):307–16.
28. **Warren JL, Yabroff KR, Meekins A, Topor M, Lamont EB, Brown ML.** Evaluation of trends in the cost of initial cancer treatment. J Natl Cancer Inst. 2008;100(12):888–97.
29. **Froom P, Melamed S, Kristal-Boneh E, Benbassat J, Ribak J.** Healthy volunteer effect in industrial workers. J Clin Epidemiol. 1999;52(8):731–5.
30. **Heilbrun LK, Nomura A, Stemmermann GN.** The effects of nonresponse in a prospective study of cancer: 15-year follow-up. Int J Epidemiol. 1991;20(2):328–38.
31. **Peppercorn J, Partridge A, Burstein HJ, Winer EP.** Standards for follow-up care of patents with breast cancer. Breast. 2005;14(6):500–8.
32. **Ahern TP, Lash TL, Thwin SS, Silliman RA.** Impact of acquired comorbidities on all-cause mortality rates among older breast cancer survivors. Med Care. 2009;47(1):73–9.
33. **Nagel G, Wedding U, Rohrig B, Katenkamp D.** The impact of comorbidity on the survival of postmenopausal women with breast cancer. J Cancer Res Clin Oncol. 2004;130(11):664–70.
34. **Duffy CM, Clark MA, Allsworth JE.** Health maintenance and screening in breast cancer survivors in the United States. Cancer Detect Prev. 2006;30(1):52–7.
35. **Snyder CF, Frick KD, Kantsiper ME, et al.** Prevention, screening, and surveillance care for breast cancer survivors compared with controls: changes from 1998 to 2002. J Clin Oncol. Jan 21 2009.
36. **Snyder CF, Frick KD, Peairs KS, et al.** Comparing care for breast cancer survivors to non-cancer controls: a five-year longitudinal study. J Gen Intern Med. Jan 21 2009.
37. **Earle CC, Neville BA.** Under use of necessary care among cancer survivors. Cancer. 2004;101(8):1712–9.
38. **Ries L, Melbert D, Krapcho M, et al.** SEER Cancer Statistics Review, 1975–2004. Bethesda, MD; 2007.
39. **Ananthakrishnan AN, Schellhase KG, Sparapani RA, Laud PW, Neuner JM.** Disparities in colon cancer screening in the Medicare population. Arch Intern Med. 2007;167(3):258–64.

Health Serv Outcomes Res Method (2010) 10:16–32
DOI 10.1007/s10742-010-0063-1

Surgery volume, quality of care and operative mortality in coronary artery bypass graft surgery: a re-examination using fixed-effects regression

Amresh D. Hanchate · Therese A. Stukel · John D. Birkmeyer · Arlene S. Ash

Received: 27 September 2007 / Revised: 26 July 2010 / Accepted: 28 July 2010 /
Published online: 10 August 2010

Abstract For many surgical procedures, apparent volume–outcome relationships may reflect differences in patient risk-profiles as well as quality of care. As some important patient profile differences may be unobserved, we use fixed effects (FE) regression to estimate the relationship between operative mortality and surgeon and hospital volumes, and compare this method with the more commonly used random effects (RE) regression approach. The 1998 and 1999 Medicare Inpatient and Denominator files for Medicare Fee for Service enrollees aged 65–99. Operative mortality rates are estimated for different surgeon and hospital volume tertiles (high, medium, low) using FE and RE regression methods, adjusted for patient demographics and morbidities. The data were collected by the Centers for Medicare and Medicaid Services (CMS). FE regression estimates that lowest volume tertile hospitals have 1.4 and lowest volume tertile surgeons have 1.6 additional

A. D. Hanchate (✉)
Section of General Internal Medicine, Boston University School of Medicine,
801 Massachusetts Ave #2077, Boston, MA 01532, USA
e-mail: hanchate@bu.edu

A. D. Hanchate
COLMR, VA Boston Healthcare System, Boston, MA, USA

T. A. Stukel
Institute for Clinical Evaluative Sciences, Toronto, ON, Canada

T. A. Stukel
Department of Health Policy, Management and Evaluation, University of Toronto & Biostatistics and Health Policy, Toronto, ON, Canada

T. A. Stukel
The Dartmouth Institute for Health Policy and Clinical Practice, Hanover, NH, USA

J. D. Birkmeyer
Department of Surgery, University of Michigan, Ann Arbor, MI, USA

A. S. Ash
Department of Quantitative Health Sciences, University of Massachusetts Medical School, Worcester, MA, USA

operative deaths (for every 100 CABG surgeries) compared to their highest volume tertile counterparts. The corresponding RE estimates are 0.5 and 1.4 respectively. The substantially higher FE hospital volume effect compared to RE indicates the presence of unobserved "protective" characteristics in lower volume providers, including a less complicated patient profile. Lower hospital and surgeon volumes are associated with substantially higher excess operative mortality from CABG surgeries than previously estimated.

Keywords Hospital volume · Surgeon volume · Fixed effects · Random effects

1 Introduction

An overwhelmingly large proportion of studies to date indicate that hospitals and surgeons who perform more surgeries have lower operative mortality (Halm et al. 2002; Shahian and Normand 2003). To what extent then can we infer that higher volume providers offer better quality care? At the provider level, lower operative mortality can result from better quality of care or from having more patients with fewer medical complications. Such variation of patients across providers can result from selective physician referral, patient choice or strategic selection by providers ("cherry picking"). An analyst using administrative or clinical data can identify and control for some of these patient profile differences using available clinical and demographic information. But there may still be other patient characteristics, largely unobserved by the analyst, that potentially influence operative mortality. We use an econometric method—"fixed effects" (FE) regression—that takes advantage of the dual clustering of patients among surgeons and hospitals, to estimate the association between provider volume and operative mortality in a way that adjusts even for unobserved provider-level differences in patient characteristics. Results are then compared with those from random effects (RE) regressions, the most commonly used approach in the literature.[1] Also referred to as 'hierarchical' models, they are commonly used in risk adjustment for evaluating health outcomes (Christiansen and Morris 1997; Krumholz et al. 2006; Normand et al. 1997; O'Brien et al. 2008; Shahian et al. 2005).

While the present study is limited to CABG surgeries, the significance of this methodological issue should be viewed in the context of the far-reaching impact of this voluminous literature, dating back over three decades (Luft et al. 1979). As indicated by recent surveys, interest in the volume–mortality relationship spans a wide range of surgeries (predominantly high risk cardiovascular surgeries and cancer resections) as well as non-surgical care (such as inpatient treatment for pneumonia or HIV positive patients)(Halm et al. 2000; Post et al. 2010; Shahian and Normand 2003). The findings from this literature

[1] As the term "fixed effects" used in the econometrics literature is quite different from that in the applied statistics/biostatistics literature, a clarification is in order. In regressions involving multilevel or hierarchical data, fixed effects regression is an approach in the econometrics literature wherein only intra-level variation is used to obtain regression coefficient estimates. For instance, in the context of longitudinal person-level data, fixed effects regression estimates are obtained by comparing only the temporal changes in measures, not the levels between individuals. The strength of this approach is that each individual is treated as a control for himself/herself. Random effects estimates utilize both within- and between-individual variation. (Johnston and DiNardo 1997; Wooldridge 2002). Alternately, in the applied statistics literature, the terms "fixed effects" and "random effects" are used to specify the nature of the regression coefficients in a multilevel regression. If a coefficient is permitted to vary—say across individuals in a longitudinal data framework—then it could be specified as a random variable; however, if a coefficient is not permitted to vary then it is described as a "fixed effect" (Raudenbush and Bryk 2002). Such a distinction does not arise in this study as none of the regression coefficients are random.

have attracted the attention of a variety of interest groups—consumer advocacy groups, health insurance coalitions and state agencies seeking to reduce costs and improve quality by enforcing protocols of proven efficacy. Strategies include regionalizing selected surgeries, publishing provider report cards and recommending provider minimum volumes for specified surgeries (Birkmeyer 2000, 2004; Birkmeyer and Birkmeyer 2006; Fredenheim 2005; Hewitt 2000).

Along with the growth of the volume-mortality literature has come a better understanding of its vulnerabilities, especially given that virtually all such studies are based on observational data, usually administrative or clinical databases. An enduring potential weakness arises from unobserved differences among patients seen by different providers. The multitude of processes that connect patients with surgeons and hospitals are not random, and are known to result in systematic differences, often large, in the patients treated by different surgeons at different hospitals (Auerbach et al. 2009; Kumbhani et al. 2009). Sicker patients are likely to gain more from higher-quality care and thus may be more prevalent among the patients of providers perceived to have higher quality. This matching may result from physician referrals or from patients' choice based on available information (including, report cards). Another sorting process involves a different kind of response to provider report cards: providers avoiding sicker patients ('cherry picking') possibly seeking a more favorable operative mortality record (Dranove et al. 2003). Recent evidence also points to systematic differences by hospitals in socioeconomic patient profiles—in particular, of significant clustering of black and other minority patients in relatively few hospitals (Jha et al. 2007; Losina et al. 2007). Not all the important differences among patients are observed, even in detailed clinical databases (Dranove et al. 2003; Gowrisankaran and Town 1999). To the extent that these unobserved factors significantly affect operative mortality, then traditional comparisons of operative mortality rates across all surgeons and hospitals risk mistaking differences in patient severity with differences in the quality of care. For instance, if higher volume hospitals attract disproportionately larger number of sicker patients and some important illness characteristics are not observed, then RE regression is likely to over-estimate adjusted mortality and under-estimate quality of care for the surgeons in higher volume hospitals. An attractive alternative is to limit comparisons of surgeons within each hospital, thereby sweeping out unobserved patient characteristics across hospitals. This is the basic logic behind the econometric approach to FE regression.

Given the clustering of patients at the surgeon and hospital levels, the advantage of FE regression is in exploiting the within-provider variation in operative mortality to estimate the volume–outcome relationships, thereby making it robust to systematic differences in unobserved characteristics at provider level. In contrast, the RE regression, the standard workhorse in this literature, is based on the assumption that there are no systematic differences across providers in unobserved patient characteristics. Given evidence of large socioeconomic patient profile differences across hospitals (due to residential segregation and other factors), cherry-picking and patient choice, the assumption of a standard random effects model needs to be validated. The only existing study using FE regression that we know of examined associations of longitudinal changes in hospital volume on two patient outcomes (length of stay and inpatient mortality) from hip fracture surgery (Hamilton and Ho 1998). They found that the apparent protective effect of higher volume from RE regression disappeared when adjusted for hospital fixed effects.

We use a readily available data set previously used to examine the association of operative mortality with hospital and surgeon volumes. To control for systematic provider-level differences in unobserved patient as well as provider characteristics, we use a FE regression approach modified to take advantage of the nature of surgeon and hospital level

clustering of patients. Two separate FE regressions are estimated, one to estimate the surgeon volume effect (i.e., association of surgeon volume and operative mortality) and another to estimate the hospital volume effect. As most hospitals in our data had two or more surgeons, to estimate surgeon volume effect, unobserved differences in patients across hospitals are controlled for by only comparing operative mortality of surgeons within same hospital. To estimate the hospital volume effect, we take advantage of the fact that surgeons are not nested in hospitals—a large proportion (51%) of surgeries are performed by surgeons who operated at two or more hospitals. This enables us to compare operative mortality across hospitals of patients operated by the same surgeon—thereby adjusting for systematic unobserved patient characteristics at the surgeon level.

By relying only on within-cluster comparisons, the FE approach offers a better approximation of the operative mortality differences arising from differences in quality of care indicators—including caregiver skill, experience and pre and post operative process of care. This is herein referred to as the *quality of care component* of the operative mortality differences by provider volume. Based on this we can also estimate the *unobserved factors component*—the residual mortality differences by provider volume that may be attributed to unobserved characteristics, including systematic unobserved patient differences across providers. We compare the estimates of this FE decomposition with the overall single estimate from a parallel RE regression. That is, does the sum of the two FE components equal the estimate from RE regression?

2 Data and methods

2.1 Data

We use an analytic data set of patient level CABG surgery mortality outcomes and covariates previously used to examine volume–outcome relationships, adopting all the variable definitions from that study (Birkmeyer et al. 2002, 2003). Briefly, using 100% of acute care hospitalization discharge data for Medicare fee-for-service beneficiaries in 1998 and 1999, admissions in which CABG surgery was performed for persons aged 65–99 are included (thereby excluding a small number of CABGs performed on younger patients with disability or End Stage Renal Disease). Discharges that also involved a valve replacement were excluded. The Institutional Review Board at Boston University School of Medicine approved the study protocol.

To identify the operating surgeon for each CABG, the unique provider identification number in the "primary operator" field in the Medicare Inpatient file was used. In 6% of procedures the provider identification numbers were invalid and therefore excluded. In addition, only CABGs performed by self-designated cardiothoracic surgeons were selected, to avoid cardiologists being wrongly identified as surgeons. This results in an additional 13% of the records being excluded, leading to a study sample of 220,592 patients.

2.2 Analytic cohorts

Patients in the data are clustered at the level of surgeons and hospitals. The FE approach to estimating the effect of surgeon volume consists of comparing surgeons in each hospital—thereby requiring at least two surgeons in every hospital. This cohort, herein called the *Within Hospital Cohort*, is obtained by excluding 60 hospitals (out of 958), because only one surgeon operated there, and 2,802 patients (out of 220,592).

An analogous FE approach is used to estimate the effect of hospital volume. Here we limit our analysis to outcome for patients of surgeons who operate at two or more hospitals ("splitters") so that we can compare outcome of patients from the same surgeon but at different hospitals. This cohort, called the *Within Surgeon Cohort*, retains 44% of all surgeons, 79% of the hospitals and 51% of all patients.

2.3 Outcome measure and covariates

Operative mortality for a patient is defined as death within 30 days of the procedure or before hospital discharge. Surgeon and hospital volumes are defined as the total number of CABG surgeries performed in a year, including Medicare as well as other payer patients. These are estimated for individual surgeons and hospitals by scaling up Medicare FFS volume to reflect total volume—the scale up multipliers are based on the proportion of all CABG patients who are Medicare FFS beneficiaries, obtained from 1997 Nationwide Inpatient Sample (NIS) and urban/rural location. Both surgeons and hospitals are categorized into patient-level tertiles—using 101 (33rd percentile) and 162 (66th percentile) surgeries per year as the cutoffs for surgeon volume and 314 and 628 for hospital volume. Patient covariates include age, sex, race, the Charlson comorbidity score and an area-level income measure (mean income from Social Security for the patient's residence ZIP code). The Charlson score is based on ICD-9-CM diagnosis codes from the index admission as well as any other inpatient admissions in the preceding 6 months, excluding primary indicators for the surgical procedure or post-operative complications. Hospital characteristics adjusted for are teaching status and ownership (not-for-profit, government and for-profit).

2.4 Estimation

Our interest is in estimating the relationship between operative mortality and both surgeon and hospital volumes. A general regression notation that encompasses the FE and RE approaches is as follows.

$$OM_{psh} = \alpha * PC_{psh} + \beta_S * SV_{sh} + \beta_H * HV_{sh} + \gamma * HC_{sh} + u_s + v_h + e_{sph} \qquad (1)$$

We reiterate that the terms random and fixed effects are not to be interpreted as describing whether regression coefficients are random variables or not; as described earlier, they refer to whether coefficients are estimated using only within-level variation (fixed effects) or within- and between-level variation (random effects). *OM* denotes operative mortality (with values 0 and 1, denoting survival and death respectively) of patient p operated by surgeon s in hospital h. The covariates are grouped as patient characteristics (*PC*), surgeon volume (*SV*), hospital volume (*HV*) and other hospital characteristics (*HC*). Unobserved cluster effects at the surgeon and hospital levels are denoted by u and v respectively. Finally e denotes the effect of unobserved patient characteristics. Both FE and RE regression models are estimated as linear regressions.[2] In both models e is specified to be independent and identically (normal) distributed random variable with mean 0. Mean of u and v are specified to be 0. Since all regression covariates are categorical groups, all

[2] Fixed effects logistic regression (also known as conditional logistic regression) requires at least one decedent and survivor from each fixed level (surgeon, hospital) (Chamberlain 1980). As 507 of the 2772 surgeons have no decedents, we have instead followed previous studies (Tsai et al. 2006) and chosen to use the linear probability specification that has the advantage of retaining data from all surgeons and hospitals.

the regression coefficients are interpreted as excess rates in operative mortality compared to the reference category.

RE estimates are obtained from a three-tiered hierarchical regression wherein surgeons are treated as being nested with in hospitals—with surgeons operating at two or more hospitals treated as distinct surgeons, using their combined volume across all hospitals (Raudenbush and Bryk 2002). This regression is estimated for both the analytic cohorts (within hospital and within surgeon) using the *xtmixed* procedure in Stata 9.2 (StataCorp 2005).

The FE estimates are obtained from two separate regressions, each estimating the volume effects for the two provider types, surgeons and hospitals. To estimate the effect of surgeon volume we limit comparisons of surgeons within the same hospital by the following transformation of Eq. 1, wherein the outcome measure as well as all the covariates are expressed in terms of within-hospital mean differences (Johnston and DiNardo 1997).

$$(OM_{psh} - \overline{OM}_h) = \alpha * (PC_{psh} - \overline{PC}_h) + \beta_s * (SV_{sh} - \overline{SV}_h) + (u_s - \overline{u}_h) + (e_{sph} - \overline{e}_h) \quad (2)$$

This transformation of all measures in terms of intra-cluster differences is the hallmark of the econometric fixed effects regression. Each variable transformation involves differencing cluster level means—for instance, $\overline{OM}_h$ denotes the mean operative mortality among all patients at hospital h. And $\overline{u}_h$ denotes the mean unobserved surgeon effects (u_s) across all the surgeons in hospital h. Note that with all hospital measures (*HV*, *HC* and *v*) eliminated, the resulting Eq. 2 has a two-tier structure (patients and surgeons). This transformed equation is estimated as a RE linear regression model using the within-hospital cohort—therefore, our approach is not purely fixed effects; instead it is a partial or quasi FE approach. The RE regression of (2) assumes that the mean differenced unobserved surgeon effect, $(u_s - \overline{u}_h)$, is uncorrelated with other covariates and error term, all of which are mean differenced—implying that, for instance, within-hospital differences around mean in volume across surgeons, $(SV_{sk} - \overline{SV}_h)$, is uncorrelated with within-hospital differences around mean in patient severity across surgeons, $(u_s - \overline{u}_h)$. However, this permits cluster-level measures (say, $\overline{SV}$ and $\overline{u}_h$) to be correlated.

Analogously, to estimate the effect of hospital volume we use the within-surgeon cohort with only patients of surgeons who operated at two or more hospitals—thereby permitting comparison of outcomes of patients across hospitals holding the surgeon characteristics the same. The corresponding transformation of (1) is

$$\begin{aligned}(OM_{psh} - \overline{OM}_s) = \alpha * (PC_{psh} - \overline{PC}_s) + \beta_h * (HV_{sh} - \overline{HV}_s) + \gamma(HC_{sh} - \overline{HC}_s) \\ + (v_h - \overline{v}_s) + (e_{sph} - \overline{e}_s)\end{aligned} \quad (3)$$

$\overline{OM}_s$ denotes the mean operative mortality for surgeon s patients across all the hospitals. Note that here surgeon characteristics (*SV* and *u*) are eliminated and the resulting structure is two-tiered (patients and hospitals). This equation is estimated as a RE linear regression model using the within-surgeon cohort.

The FE hospital volume effect is analogously estimated by limiting comparisons of operative mortality of patients with the same surgeon but who operates at different hospitals (i.e., the within surgeon cohort). Here transformation involves differencing surgeon-level averages, followed by a two-tiered RE regression involving hospital unobserved cluster effect. Additional technical details are elaborated in the Appendix. The *quality of care component* of the volume effect on operative mortality is measured by β_H for hospital volume and β_S for surgeon volume, both estimated from the FE regressions.

FE and RE approaches differ principally in the specification of cluster effects u_s and v_h. Note that v_h represents the operative mortality at hospital h resulting from unobserved factors that vary systematically across hospitals, and u_s is the analog for unobserved differences across surgeons. Therefore systematic differences in unobserved patient characteristics, if any, are captured by either u_s and v_h. The RE estimation assumes that u_s and v_h are uncorrelated with the model covariates (*PC, SV, HV* and *HC*) as well as residual e—in particular, this implies that there are no systematic differences in unobserved patient characteristics across providers. Violation of this assumption results in biased estimates—β_H and β_S from RE no longer reflect the *quality of care component*. In contrast the FE approach makes a weaker assumption, permitting this correlation at both the cluster levels (surgeons and hospitals), but assuming that within-cluster differences across measures are uncorrelated. This implies that even if provider volume were correlated with unobserved patient severity, β_H and β_S from FE remain unbiased estimates of *quality of care component*. Therefore, the greater the differences in FE and RE estimates of β_H and β_S, the greater is the influence of unobserved factor differences by providers.

The overall *unobserved factors component* is measured by the residual mortality—that is, the combined effect of unobserved hospital and surgeon factors ($u + v$). This is obtained by substituting the FE covariate estimates in Eq. 1 to obtain

$$OM_{psh} - (\alpha * PC_{psh} + \beta_S * SV_{sh} + \beta_H * HV_{sh} + \gamma * HC_{sh}) = u_s + v_h + e_{sph} \qquad (4)$$

We report the average of this measure by each provider volume.

Therefore, we not only compare RE estimates of β_H and β_S against those from FE, but also against the sum of *quality of care* and *unobserved factors components* from FE regressions.

3 Results

3.1 CABG patients, surgeons and hospitals

Table 1 provides sample characteristics of the entire study population (All column) as well as the two subsets—the within hospital and the within surgeon cohorts. The entire study population consists of 220,592 patients operated on by 2,772 surgeons in 958 hospitals during 1998 and 1999. The overall operative mortality was 50.5 per 1,000 surgeries. The average annual volume for surgeons was 85 and that for hospitals 297—note that these volumes refer to not only Medicare patients but to all patients. As mentioned earlier, 44% of surgeons operated at two or more hospitals (i.e., splitters) and 94% of hospitals had two or more surgeons.

All analyses are performed in terms of patient-level tertiles of hospital and surgeon volumes. Among the 2,772 surgeons, 377 were in the top tertile, with annual volume ranging from 162 to 567 surgeries, and 1,783 were in the bottom tertile, with a volume of less than 101 surgeries. Of the 958 hospitals, 101 are in the top tertile (again of patients) with an annual volume of at least 628 surgeries; 644 hospitals are in the bottom tertile with a volume of less than 314 surgeries. Patients in high/low volume hospital (i.e., highest/lowest volume tertile) were more likely to also have a high/low volume surgeon, and vice versa. Of the patients at high volume hospitals, half were operated on by high volume surgeons and about 15% by low volume surgeons; for the patients at low volume hospitals, half were operated on by low volume surgeons and about 15% by high volume surgeons. A

Table 1 Patient, surgeon and hospital characteristics: CABG admissions, 1998–99

	All	Within hospital cohort[a]	Within surgeon cohort[a]
# CABG admissions[b]	220,592	217,790	112,143
Mean operative mortality rate (per 1,000)	50.5	50.4	52.1
# Surgeons	2,772	2,744	1,216
Mean annual # CABG surgeries per surgeon	85	86	96
% Surgeons operating at 2 or more hospitals	44%	44%	100%
# Hospitals	958	898	755
Mean annual # CABG surgeries per hospital	297	311	250
% Hospitals with 2 or more surgeons	94%	100%	94%
Provider volume			
% Patients with surgeon volume <101 ("lowest tertile surgeon")[c]	33%	33%	33%
% Patients with surgeon volume >162 ("highest tertile surgeon")[c]	33%	33%	36%
% Patients in hospitals with volume <314 ("lowest tertile hospital")[c]	33%	33%	42%
% Patients in hospitals with volume >628 ("highest tertile hospital")[c]	33%	34%	28%
Patient characteristics (%)			
Age	40%	40%	40%
65–74	60%	60%	60%
75–84	37%	37%	37%
85+	3%	3%	3%
Female	35%	35%	35%
Black	4%	4%	4%
Charlson score[d]	10%	10%	10%
0	41%	41%	41%
1	33%	33%	34%
2	16%	16%	16%
3+	1%	1%	1%
Nonelective admission	57%	57%	56%
Resident zip code mean Social Security income below $2,500	66%	66%	67%
Hospital characteristics (% patients with following hospital characteristics)			
Teaching	46%	46%	39%
Government owned	7%	7%	5%
Not for profit	81%	81%	80%
For profit	11%	11%	15%

All patients are 65 or older and Medicare Fee for Service enrolled

[a] Within Hospital Cohort excludes CABG admissions from 60 hospitals with only one surgeon, and Within Surgeon Cohort retains CABG admissions from the 1,216 surgeons who operated at two or more hospitals

[b] Among all these admission, there are no patients with more than one CABG surgery—each surgery admission refers to a distinct patient

[c] Note that surgeon and hospital volumes are estimates of annual volumes covering all payers

[d] Note that the Charlson score, based on ICD-9 diagnoses codes from the index admission as well as any other inpatient admissions in the preceding 6 months, excludes primary indicators for the surgical procedure or post-operative complications

similar pattern was observed for the converse distribution of patients at low and high volume surgeons across low and high volume hospitals.

3.2 The analytic cohorts

All summary figures in Table 1 for the within hospital cohort are virtually identical to that for the entire sample. However, the within surgeon cohort, containing surgeons operating at two or more hospitals, shows differences in provider profiles—surgeon volumes are larger (by 10%), hospital volumes are smaller (by 17%) and fewer hospitals are teaching (by 15%). More importantly, the patient profile appears to be no different compared to the overall population, and operative mortality rates and patient characteristics are very similar.

Table 2 gives the unadjusted operative mortality rates (per 1,000 CABG surgeries) for different strata cross-classified by provider volume for both analytic cohorts. Lower volume hospitals had 11 more operative deaths while lower volume surgeons had 14 more operative deaths compared to their high volume counterparts. Higher mortality is associated with older age, female gender, black race, higher Charlson score and emergent admission. With respect to provider volume two patterns emerged—(i) the magnitude of difference between high and low volume providers (either hospitals or surgeons) is constant across most strata, and (ii) the two analytic cohorts have very similar operative mortality rates for the same strata.

3.3 Estimates of surgeon and hospital volume effects

Table 3 presents the main regression estimates from both FE and RE approaches. Column 1 presents the RE estimates of operative mortality rates (%) associated with surgeon and hospital volumes using the within hospital cohort. For comparison this model is also estimated for the within surgeon cohort (column 2). Columns 3 and 4 give the corresponding estimates from the FE approach. Note that both the RE estimates are similar across all patient and hospital characteristics, although the hospital volume effects are smaller in column (2) but not (statistically) significantly different—the column (1) figure will be used for RE estimates herein. Using patients at high volume providers as the reference cohort, those treated by low and medium volume hospitals had 0.45 more operative deaths per 100 CABG surgeries (95% CI = [0.0008, 0.89]) and those treated by low volume surgeons had 1.41 more operative deaths (95% CI = [1.08, 1.74]). The FE approach estimates that, compared to their high volume provider counterparts, patients treated by low volume hospitals had 1.36 more operative deaths (95% CI = [0.34, 2.37] and those by low volume surgeons had 1.56 more operative deaths (95% CI = [0.67, 2.46]).

Estimates of other covariates that have statistically significant association with operative mortality—gender and age composition, Charlson score and admission type, have very similar estimates across all four regressions (columns 1–4). The teaching status of the hospital and the mean Social Security income in a patient's residence zip code were not associated with operative mortality in any of the models. However, for-profit hospitals had higher operative mortality in the RE models but not in the FE model.

As described in the Methods section, FE regression is used to decompose the total volume effect into *quality of care* component (obtained from FE regression coefficients in Table 3) and *unobserved factor* component (hospital or surgeon-level mean of FE regression residuals). These are presented in the first two columns of Table 4, followed by

Table 2 Operative Mortality Rate (%) for Highest & Lowest Volume Tertiles, by Sample Characteristics

	Within Hospital Cohort (Hospitals with at least 2 surgeons)				Within Surgeon Cohort (Surgeons practicing at 2 or more hospitals)			
	Hospital volume low (<314)	Hospital volume high (>628)	Surgeon volume low (<101)	Surgeon volume high (>162)	Hospital volume low (<314)	Hospital volume high (>628)	Surgeon volume low (<101)	Surgeon volume high (>162)
All	5.6	4.5	5.9	4.5	5.7	4.4	6.2	4.6
Age								
65–69	3.5	3.0	3.9	2.9	3.6	3.0	3.9	3.0
70–74	5.0	4.0	5.4	3.8	5.1	3.9	5.8	3.9
75–79	6.2	5.3	6.7	5.3	6.3	5.1	6.9	5.3
80–84	9.3	7.0	9.6	7.1	9.8	6.8	10.3	7.2
85+	12.1	8.5	12.9	9.0	11.5	8.2	12.7	10.3
Gender								
Female	7.0	5.6	7.4	5.4	7.2	5.6	7.9	5.5
Male	4.8	3.9	5.2	4.0	4.9	3.8	5.3	4.1
Race								
Black	6.6	5.5	6.9	5.1	6.2	4.8	6.9	4.7
Non-Black	5.5	4.5	5.9	4.4	5.7	4.4	6.2	4.6
Charlson score								
0	5.6	4.7	6.1	4.5	5.7	4.5	6.3	4.5
1	5.2	4.1	5.4	4.2	5.3	3.9	5.8	4.3
2	5.2	4.1	5.5	4.1	5.3	4.2	5.6	4.3
3+	7.3	6.2	7.9	5.8	7.5	6.2	8.3	6.1
Resident zip code mean Social Security income								
Below $2,500	5.6	4.6	6.1	4.6	5.8	4.5	6.4	4.6
Above $2,500	5.5	4.3	5.7	4.3	5.5	4.2	5.7	4.5
Admission type								
Elective	4.2	3.4	4.4	3.3	4.5	3.4	4.7	3.5
Urgent/ emergent	6.6	5.4	7.1	5.4	6.6	5.2	7.4	5.5
Teaching status								
Teaching hospital	5.6	4.5	5.8	4.4	5.9	4.2	6.0	4.6
Not teaching hospital	5.5	4.5	6.1	4.5	5.6	4.7	6.3	4.6
Hospital ownership								
Government owned	5.6	5.3	6.0	5.1	6.2	6.5	6.5	5.4
Not for profit	5.4	4.5	5.8	4.4	5.5	4.4	6.0	4.5
For profit	6.0	3.8	6.6	4.6	6.1	3.1	6.8	4.8

their sum in column 3. As the *unobserved factor* component (column 2) gives the portion of the observed operative mortality not explained by the regression variables, negative values indicate that observed mortality was less than expected by the FE model. We see

Table 3 Random and fixed effects regression estimates of excess operative mortality (per 100 CABG surgeries)

	Random Effects				Fixed Effects			
	Within hospital cohort (1)		Within surgeon cohort (2)		Within hospital cohort (3)		Within surgeon cohort (4)	
	Co-eff.	Std. Err.	Co-eff.	Std. Err.	Co-eff.	Std. Err.	Co-eff.	Std. Err.
Hospital volume								
Lowest tertile (<314)	**0.45**	0.23	0.09	0.30			**1.36**	0.52
Middle tertile (314–628)	0.39	0.23	0.16	0.28			0.80	0.53
Highest tertile (>628)	Reference		Reference				Reference	
Surgeon volume								
Lowest tertile (<101)	**1.41**	0.17	**1.53**	0.26	**1.56**	0.46		
Middle tertile (101–162)	0.24	0.16	0.27	0.25	−0.11	0.47		
Highest tertile (>162)	Reference		Reference		Reference			
Female	**1.43**	0.10	**1.21**	0.14	**1.43**	0.10	**1.71**	0.14
Age								
65–69	Reference		Reference		Reference		Reference	
70–74	**1.12**	0.12	**1.06**	0.17	**1.13**	0.12	**1.22**	0.17
75–79	**2.34**	0.13	**2.29**	0.18	**2.35**	0.13	**2.40**	0.18
80–84	**4.51**	0.16	**4.21**	0.23	**4.51**	0.16	**4.83**	0.23
85+	**7.05**	0.30	**7.03**	0.42	**7.01**	0.30	**7.16**	0.42
Black	**0.31**	0.25	**0.77**	0.36	**0.07**	0.26	**−0.23**	0.36
Charlson score								
0	Reference		Reference		Reference		Reference	
1	**−0.52**	0.11	**−0.67**	0.15	**−0.52**	0.11	**−0.37**	0.15
2	**−0.33**	0.14	**−0.40**	0.19	**−0.31**	0.14	**−0.24**	0.20
3+	**1.75**	0.17	**1.54**	0.23	**1.77**	0.17	**2.04**	0.24
Admission type								
Elective	Reference		Reference		Reference		Reference	
Urgent/emergent	**2.16**	0.10	**2.19**	0.14	**2.25**	0.10	**2.14**	0.14
Year								
1998	Reference		Reference		Reference		Reference	
1999	0.03	0.09	−0.03	0.13	0.08	0.10	0.06	0.14
Major teaching hospital	−0.13	0.15	−0.19	0.21			0.05	0.28
Hospital ownership								
Not for profit	Reference		Reference				Reference	
Government	0.38	0.28	0.23	0.34			0.14	0.77
For profit	0.61	0.24	0.66	0.41			−0.16	0.45
Resident zip code mean Social Security income below $2,500	0.16	0.11	0.18	0.15	0.05	0.11	0.05	0.16

Note: Estimates in bold are significant at 5% level

that low volume hospitals have 0.84 fewer deaths than expected (using highest volume tertile as the reference); and middle (tertile) volume hospitals have 0.40 fewer deaths—resulting from unobserved factors including a less complicated patient profile. Column 3

Table 4 Expected excess operative mortality rates (%) by hospital & surgeon volume

	(1) Quality of care component[a] (FE expected)	(2) Unobserved factor component (FE residual)	(3) Total FE (1) + (2)	(4) RE[a]
Hospital volume				
Lowest tertile (<314)	1.36	−0.84	0.52	0.45
Middle tertile (314–628)	0.80	−0.40	0.40	0.39
Highest tertile (>628)	Reference	Reference	Reference	Reference
Surgeon volume				
Lowest tertile (<101)	1.56	−0.58	0.98	1.41
Middle tertile (101–162)	−0.11	0.07	−0.04	0.24
Highest tertile (>162)	Reference	Reference	Reference	Reference

[a] These figures are from Table 3 (FE Expected values are from columns 3 or 4 while RE values are from column 1)

gives the sum of the two FE components. Combining the two components—1.36 additional deaths from the quality component and 0.84 fewer deaths due to other unobserved factors, lowest tertile hospitals have a total (net) excess mortality of 0.52 deaths (per 100 CABG surgeries)—as indicated in column 3. This figure is similar to the direct estimate of 0.45 excess deaths from RE regression (column 4). A similar pattern is true for medium volume hospitals, with a net FE effect of 0.40 and a direct RE estimate is 0.39 excess deaths, thereby suggesting that RE hospital volume effects approximate the sum of the quality and unobserved factor FE components. But this comparison does not hold for the surgeon volume comparison—particularly for the low volume surgeons. The average *unobserved factors component* for low volume surgeons is −5.8, thereby leading to the total FE estimate of 9.8 excess mortality rate—much lower than the RE direct estimate of 14.1.

3.4 Sensitivity analyses

An important step in the estimation process is the use of the two distinct subsets (cohorts) of the overall data. In particular, the within surgeon cohort excludes 49% of the patients and 44% of surgeons. We performed a number of robustness and sensitivity checks to validate the estimates obtained. First, to assess the robustness of the FE hospital effects estimates, we re-estimated the regression using bootstrapping (1,000 replications) randomly dropping 10% of the surgeons (and their patients). The results do not change—the average excess mortality rate for low volume hospitals is 1.4 (95% CI = [0.89, 1.86]) and for medium volume hospitals is 0.8 (95% CI = [0.42, 1.29]).

Second, we address the difference in hospital volume effects estimated by the two RE regressions (columns 1 and 2 of Table 3). For instance, for low volume hospitals, the excess mortality estimate is 0.45 using the within hospital cohort (column 1) but it is 0.09 using the within surgeon cohort (column 2). Note that while the within hospital cohort includes virtually all the study data, the within surgeon cohort excludes all patients treated by 44% of surgeons who operated at more than one hospital. It is unclear if this difference is indicative of systematic difference in the within surgeon cohort subset or indicative of lack of robustness of the estimate. To evaluate this we obtained 100 different sub-samples

of the overall data by excluding all patients from 44% of surgeons randomly selected—that is, exclusion of surgeons was not based on whether or not they operated at more than one hospital. The mean (from the 100 regressions) of low hospital volume effect is 0.39, with the range [−0.17, 0.88]—thereby suggesting that the aforementioned difference (betweens columns 1 and 2 of Table 3) may be due to the lack of robustness of the estimate and not necessarily indicative of systematic differences in patients in the two cohorts.

To further ascertain if the patients included in the within surgeon cohort are systematically different from those who are excluded, and in the spirit of marginal propensity score estimation, we performed a logistic regression with inclusion/exclusion as the outcome and all patient factors plus operative mortality as the covariates. We find that this model has poor discrimination—with only 51% of patients correctly classified (area under ROC was 51.3), thereby indicating little systematic difference between the included and excluded patients.

Finally, to examine if the results are sensitive to using distinct cohorts for the two FE regressions, we identified a sub-sample that simultaneously met the two criteria identifying each cohort (hospitals with at least two surgeons, and surgeons who operated at two or more hospitals)—this sub-sample has 104,340 patients, 596 hospitals and 1,126 surgeons. The RE and both the FE regressions estimated on this common data yielded virtually the same results as those in Table 3—in particular, a much higher estimate of excess deaths associated with low-hospital volume from FE regression (1.21 deaths) than from RE regression (0.66 deaths).

4 Discussion

Are lower operative mortality rates among higher volume surgeons and hospitals a signal of better quality of care, or is this association confounded by patient profile differences across providers? Providers may differ in the proportion of complicated patients they treat, and not all of these complications are adequately identifiable in administrative or clinical data. To better adjust for such differences, we used fixed effects (FE) regression methods to obtain estimates based only on within cluster comparisons. To avoid comparisons across hospitals, surgeons within same hospital were compared with each other in estimating the surgeon volume effect; and to estimate the hospital volume effect we compared operative mortality across hospitals for patients treated by the same surgeon. By overcoming the potential confounding from unobserved patient differences, this approach better approximates the volume effects associated with *quality of care* differences across providers. This regression approach indicates that, compared to hospitals that perform at least 628 CABG surgeries a year, those that perform less than 314 a year have an excess operative mortality rate of 1.36 deaths (per 100 surgeries) and those that perform between 314 and 628 surgeries have an excess mortality of 0.8. Further, compared to surgeons who perform at least 162 CABG surgeries a year, those who perform fewer than 101 surgeries have an excess operative mortality rate of 1.56.

Comparing FE and RE regression estimates indicate that they differ significantly with respect to the effects of hospital volume but not surgeon volume. Specifically, FE regression estimates that excess deaths associated with low volume hospitals (1.36 excess deaths per 100 CABG surgeries) is much higher than that estimated by RE regression (0.52). Recall that the FE estimate is based on comparing operative mortality of the same surgeon across hospitals—implying that even after adjusting for surgeon-level factors there are significant differences across hospitals in unobserved factors. On the other hand FE and

RE estimates of excess deaths associated with low surgeon volume are similar (1.56 and 1.41 respectively)—since FE estimates are based on within hospital comparison of surgeons, similarity between FE and RE estimates implies that, once unobserved factors at the hospital level are adjusted for (i.e., within each hospital), unobserved factors across surgeons by volume do not have significant effect on operative mortality. That is, within each hospital, there are no systematic differences in unobserved patient characteristics across surgeons—consequently the surgeon volume effects from RE and FE are very similar.

The FE approach also enables estimation of the effect of unobserved characteristics on operative mortality (*unobserved factors component)* at the provider level. This captures factors that are important determinants of operative mortality but not observed in the data. Averaging this component at provider volume level, we indeed find large differences in this component. Low volume hospitals and surgeons have much lower operative mortality than expected by the FE regression model. While the regression model predicts low volume hospitals to have an excess operative mortality rate of 1.36, the observed excess mortality rate is 0.52, leading to a large difference accounted for by unobserved factors. Similarly, while the low volume surgeons have an expected excess mortality rate of 1.56, the observed excess mortality rate is 0.98. Both these indicate the presence of large "protective" factors in low volume hospitals and providers. Note that process-of-care differences across providers that vary with volume are already captured explicitly in the regression, but other process differences (that affect quality) could contribute to this unobserved component.

Based on growing evidence from other studies, a plausible "protective" factor may be unobserved patient severity or complications, implying that high volume hospitals and surgeons have more complicated patients compared to their low volume counterparts. While this implication of a more complicated patient profile in higher volume hospitals contradicts some previous findings (Shahian and Normand 2003), a number of recent studies corroborate it. The strongest evidence comes from studies of the impact of mandatory annual reporting of risk-adjusted mortality from CABG surgery for each hospital and surgeon in New York and Pennsylvania since 1992. One study surveyed a sample of cardiologists and cardiac surgeons from the two states, finding that 59% of the cardiologists "reported increased difficulty in finding surgeons willing to perform CABG surgery in severely ill patients who required it, and 63% of the cardiac surgeons reported that they were less willing to operate on such patients" (Schneider and Epstein 1996). Another large study based on the majority of all CABG surgeries performed among Medicare population between 1987 and 1994, found that "report cards led to increased sorting of patients to providers on the basis of the severity of their illness … with those two states' teaching hospitals picking up an increasing share of patients with more severe illness" (Dranove et al. 2003). Since physician referral has been found to be the most important determinant of provider choice, it is likely that more complicated patients are referred to teaching and other high volume providers as they are expected to gain more than those with fewer complications (Schwartz et al. 2005). Despite this supportive evidence in the literature, we acknowledge that the present study provides not direct evidence of unobserved patient profile differences by provider volume. There may be other unobserved factors, including at provider levels, not adequately captured by the methods used here.

This study also has implications for the random effects (RE) regression approach to estimating volume effects. As this method does not adjust for unobserved factors that affect operative mortality, including important patient severity indicators, RE volume effect estimates may be an inappropriate measure of quality of care differences. Therefore if our interest is in volume effects driven by quality of care differences, say for report cards, then the appropriate estimates are those from FE regression. But if our interest is in the overall association of operative mortality with surgeon volume—either arising from quality of care

or other differences, then RE volume effect estimates appear to approximate the combined FE volume effect. Note that it is this overall measure is the appropriate measure in assessing the impact of regionalization of CABG surgery—since the quality of care benefit from regional centers is only experienced if complicated patients would otherwise have gone to "low-volume" providers, the appropriate accounting should therefore combine quality of care component with unobserved factor component from FE regression.

While the FE approach used here addresses some of the complexities of using observational data, a number of important limitations still remain. Firstly, given the two levels of clustering of patients (within surgeons and hospitals), the FE approach used here is a two-step procedure wherein the first step uses FE for one level of clustering while the second component is an RE regression at the other level. Although less restrictive than a three-tiered RE model, it nevertheless does not eliminate all forms of correlation of unobserved cluster effects. For instance, in comparing surgeons within each hospital it is assumed that patient profiles are similar across surgeons—to the extent that some of the patients are triaged to specific surgeons based on unobserved characteristics (including severity) the quality measure is confounded. Also, the fixed effects method does not allow interaction of unobserved hospital- and surgeon-level factors. Similarly, when comparing patient outcomes for the same surgeon but at different hospitals it is assumed that there are no systematic unobserved differences among patients across hospitals. The second limitation is that the range of clinical information to measure patients' disease burden is limited. The large unobserved mortality component estimated here might be the result of limited comorbidity information available (Charlson scores based only on ICD-9-CM codes from inpatient records for 6 months). Enriching this information, in particular, with data from clinical databases, might result in lower unobserved mortality. However it is not clear if the volume effects associated with quality of care differences will be affected with the use of richer patient information. The FE regression used to estimate hospital volume effects uses a subset that excludes 49% of the patient records in the available data. While we have performed a variety of sensitivity and robustness checks, there may still be other differences between those included and excluded. Third, we also recognize that this study does not attempt to disentangle the bi-directional relationship between volume and operative mortality—it only estimates a reduced form relationship between the two measures regardless of the underlying cause. While the "learning by doing" hypothesis posits volume as the cause, an alternative hypothesis ("selective referral") allows for quality (operative mortality rate) being the cause and volume the effect. Majority of the studies that have attempted to disentangle the causal direction, using instrumental variables regression, have found evidence for both effects at the hospital level (Farley and Ozminkowski 1992; Gaynor et al. 2005; Gowrisankaran et al. 2005; Luft 1980; Tsai et al. 2006)—we know of no studies that have modeled both surgeon and hospital volumes.

In conclusion, fixed effects (FE) regression estimates excess operative mortality from low volume hospitals to be much higher than previously estimated. This reflects differences in operative mortality arising from quality of care differences across providers by volume. It appears that the previous estimates using random effects (RE) regression captured not only quality of care differences but also other important operative mortality determinants unobserved in the data—in particular, unobserved patient complications.

Acknowledgment This study was funded by an AHRQ Grant (R03 HS015617-01A1).

References

Auerbach, A.D., Hilton, J.F., Maselli, J., Pekow, P.S., Rothberg, M.B., Lindenauer, P.K.: Shop for quality or volume? Volume, quality, and outcomes of coronary artery bypass surgery. Ann. Intern. Med. **150**(10), 696–704 (2009)

Birkmeyer, J.D.: Should we regionalize major surgery? Potential benefits and policy considerations. J. Am. Coll. Surg. **190**(3), 341–349 (2000)

Birkmeyer, J.D.: Understanding surgeon performance and improving patient outcomes. J. Clin. Oncol. **22**(14), 2765–2766 (2004)

Birkmeyer, J.D., Siewers, A.E., Finlayson, E.V.A., Stukel, T.A., Lucas, F.L., Batista, I., Welch, H.G., Wennberg, D.E.: Hospital volume and surgical mortality in the United States. N. Engl. J. Med. **346**(15), 1128–1137 (2002)

Birkmeyer, J.D., Stukel, T.A., Siewers, A.E., Goodney, P.P., Wennberg, D.E., Lucas, F.L.: Surgeon volume and operative mortality in the United States. N. Engl. J. Med. **349**(22), 2117–2127 (2003)

Birkmeyer, N.J.O., Birkmeyer, J.D.: Strategies for improving surgical quality—should payers reward excellence or effort? N. Engl. J. Med. **354**(8), 864–870 (2006)

Chamberlain, G.: Analysis of covariance with qualitative data. Rev. Econ. Stud. **47**, 225–238 (1980)

Christiansen, C. L., Morris, C. N.: Improving the statistical approach to health care provider profiling. *Ann Intern Med* **127**(8_Part_2):764–768 (1997)

Dranove, D., Kessler, D., McClellan, M., Satterthwaite, M.: Is more information better? The effects of "report cards" on health care providers. J. Polit. Econ. **111**(3), 555 (2003)

Farley, D.E., Ozminkowski, R.J.: Volume–outcome relationships and in-hospital mortality: the effect of changes in volume over time. Med. Care **30**(1), 77–94 (1992)

Fredenheim, M.: To find a doctor, mine the data. The New York Times, New York City (2005)

Gaynor, M., Seider, H., Vogt, W.B.: The volume-outcome effect, scale economies, and learning-by-doing. Am. Econ. Rev. **95**(2), 243–247 (2005)

Gowrisankaran, G., Town, R.J.: Estimating the quality of care in hospitals using instrumental variables. J. Health Econ. **18**(6), 747–767 (1999)

Gowrisankaran, G., Ho, V., Town, R.: Causality, Learning and Forgetting in Surgery. Working Paper, John M. Olin School of Business, Washington University in St. Louis. St. Louis (2005)

Halm, E.A., Lee, C., Chassin, M.R.: How is Volume Related to Quality in Health Care? A Systematic Review of the Research Literature. Institute of Medicine, Washington, D.C. (2000)

Halm, E.A., Lee, C., Chassin, M.R.: Is volume related to outcome in health care? A systematic review and methodologic critique of the literature. Ann. Intern. Med. **137**(6), 511–520 (2002)

Hamilton, B.H., Ho, V.: Does practice make perfect? Examining the relationship between hospital surgical volume and outcomes for hip fracture patients in Quebec. Med. Care **36**(6), 892–903 (1998)

Hewitt, M.: Interpreting the Volume–Outcome Relationship in the Context of Health Care Quality: Workshop Summary. Institute of Medicine, Washington, D.C. (2000)

Jha, A.K., Orav, E.J., Li, Z., Epstein, A.M.: Concentration and quality of hospitals that care for elderly black patients. Arch. Intern. Med. **167**(11), 1177–1182 (2007)

Johnston, J., DiNardo, J.: Econometric Methods, 4th edn. McGraw-Hill, New York, NY (1997)

Krumholz, H.M., Brindis, R.G., Brush, J.E., Cohen, D.J., Epstein, A.J. , Furie, K., Howard, G., Peterson, E.D., Rathore, S.S., Smith, S.C., Jr., Spertus, J.A., Wang, Y., Normand, S.L.: Standards for statistical models used for public reporting of health outcomes: an American Heart Association Scientific Statement from the Quality of Care and Outcomes Research Interdisciplinary Writing Group: co-sponsored by the Council on Epidemiology and Prevention and the Stroke Council. Endorsed by the American College of Cardiology Foundation. Circulation **113**(3):456–62 (2006)

Kumbhani, D.J., Cannon, C.P., Fonarow, G.C., Liang, L., Askari, A.T., Peacock, W.F., Peterson, E.D., Bhatt, D.L.: Association of hospital primary angioplasty volume in ST-segment elevation myocardial infarction with quality and outcomes. JAMA **302**(20), 2207–2213 (2009)

Losina, E., Wright, E.A., Kessler, C.L., Barrett, J.A., Fossel, A.H., Creel, A.H., Mahomed, N.N., Baron, J.A., Katz, J.N.: Neighborhoods matter: use of hospitals with worse outcomes following total knee replacement by patients from vulnerable populations. Arch. Intern. Med. **167**(2), 182–187 (2007)

Luft, H.S.: The relation between surgical volume and mortality: an exploration of causal factors and alternative models. Med. Care **18**(9), 940–959 (1980)

Luft, H.S., Bunker, J., Enthoven, A.: Should operations be regionalized? The empirical relation between surgical volume and mortality. N. Engl. J. Med. **301**(25), 1364–1369 (1979)

Normand, S.-L.T., Glickman, M.E., Gatsonis, C.A.: Statistical methods for profiling providers of medical care: issues and applications. J. Am. Stat. Assoc. **92**(439), 803–814 (1997)

O'Brien, S.M., DeLong, E.R., Peterson, E.D.: Impact of case volume on hospital performance assessment. Arch. Intern. Med. **168**(12), 1277–1284 (2008)
Post, P.N., Kuijpers, M., Ebels, T., Zijlstra, F.: The relation between volume and outcome of coronary interventions: a systematic review and meta-analysis. Eur Heart J (2010)
Raudenbush, S.W., Bryk, A.S.: Hierarchical Linear Models, 2nd edn. Sage Publications, Thousand Oaks, CA (2002)
Schneider, E.C., Epstein, A.M.: Influence of cardiac-surgery performance reports on referral practices and access to care—a survey of cardiovascular specialists. N. Engl. J. Med. **335**(4), 251–256 (1996)
Schwartz, L.M., Woloshin, S., Birkmeyer, J.D.: How do elderly patients decide where to go for major surgery? Telephone interview survey. BMJ **331**(7520), 821–827 (2005)
Shahian, D.M., Normand, S.-L.T.: The volume–outcome relationship: from Luft to Leapfrog. Ann. Thorac. Surg. **75**(3), 1048–1058 (2003)
Shahian, D.M., Torchiana, D.F., Shemin, R.J., Rawn, J.D., Normand, S.-L.T.: Massachusetts cardiac surgery report card: implications of statistical methodology. Ann. Thorac. Surg. **80**(6), 2106–2113 (2005)
StataCorp.: Stata Statistical Software: Release 9. StataCorp LP, College Station, TX (2005)
Tsai, A.C., Votruba, M., Bridges, J.F.P., Cebul, R.D.: Overcoming bias in estimating the volume–outcome relationship. Health Serv. Res. **41**(1), 252–264 (2006)
Wooldridge, J.M.: Econometric analysis of cross-section and panel data. The MIT Press, Cambridge, MA (2002)

Perceived Racial Discrimination in Health Care and its Association with Patients' Healthcare Experiences: Does the Measure Matter?

Leslie R. M. Hausmann, PhD; Nancy R. Kressin, PhD; Barbara H. Hanusa, PhD; Said A. Ibrahim, MD

Objectives: Examine whether three measures of perceived racial discrimination in health care detect similar rates of discrimination and show similar associations with patients' healthcare experiences.

Design: Cross-sectional observational study involving telephone surveys and medical record reviews.

Setting: Veterans Affairs Pittsburgh Healthcare System

Participants: 50 White and 50 African American veterans with diabetes

Main Outcome Measures: Three types of measures of perceived racial discrimination in health care were compared: single-item and multi-item measures assessing personal experiences of discrimination in healthcare settings, and a measure assessing general racism in the healthcare system. Associations of each measure with patient-reported problems with their medical care and receipt of recommended preventive screenings were also explored.

Results: More African American than White veterans reported perceived discrimination on all measures (personal discrimination, single-item: 42% vs 6%, $P<.001$; personal discrimination, multi-item: 42% vs 18%, $P=.01$; general racism: 74% vs 40%, $P=.001$). In the total sample, discrimination was more likely to be reported on the general racism measure than on the single-item (OR=36.53, 95% CI=7.95–167.89) or multi-item measures (OR=20.28, 95% CI=5.12–80.34) of personal discrimination. The multi-item measure of personal discrimination (OR=3.96, 95% CI=1.29–12.18) and general racism measure (OR=3.61, 95% CI=1.34–9.71) were significantly associated with patient-reported problems with their care. Receipt of recommended screenings was not associated with any of the discrimination measures.

Conclusions: Different measures of perceived racial discrimination in healthcare settings yield different rates of discrimination and show variable associations with patients' perceptions of care. (*Ethn Dis*. 2010;20:40–47)

Key Words: Perceived Discrimination, Quality of Care, Healthcare Utilization, Diabetes

From Health Services Research & Development Service, Veterans Affairs Pittsburgh Healthcare System, Center for Health Equity Research and Promotion, Pittsburgh, PA (LRMH, SAI, BHH); Health Services Research and Development Service, Veterans Affairs Boston Healthcare System, Boston, MA, Bedford Veterans Affairs Medical Center, Center for Health Quality, Outcomes and Economic Research, Bedford, MA, and the Section of General Internal Medicine, Boston University School of Medicine (NRK); Department of Psychiatry, University of Pittsburgh School of Medicine (BHH); Division of General Internal Medicine, Department of Medicine, University of Pittsburgh School of Medicine (SAI).

Address correspondence and reprint requests to Leslie R.M. Hausmann, PhD; VA Pittsburgh Healthcare System; Center for Health Equity Research and Promotion; 7180 Highland Drive (151C-H); Pittsburgh, PA 15206; 412-954-5221; 412-954-5264 (fax); leslie.hausmann@gmail.com

Introduction

More than 200 empirical studies have investigated the health implications of discrimination,[1–3] which refers to differential and negative treatment of individuals because of their membership in a particular demographic group (eg, race, sex, class).[4] Although discrimination can be based on any characteristic and may affect health even when it is not directly perceived,[5] most research has focused on the health effects of race or ethnicity-based discrimination that is directly perceived by targeted individuals.[1–3] Such research has demonstrated that perceived racial and ethnic discrimination is associated with decrements in both mental and physical health, as well as an increase in negative health behaviors (eg, cigarette smoking, alcohol use).[1,2,6–11]

The current study focuses on perceptions of racial discrimination encountered in healthcare settings, which has been the focus of a relatively small subset of studies.[7,12–26] Discrimination may arise in healthcare settings due to a number of factors.[4,27] For instance, geographic location and bureaucratic complexity of medical facilities may result in differential access to and utilization of health services for different racial and ethnic groups. Unequal healthcare delivery can also result from the ambiguous nature of clinical decisions, misunderstandings in doctor-patient communication, provider attitudes and stereotypes, and/or expectations patients have regarding clinical encounters.

Understanding discrimination in healthcare settings is particularly important for several reasons. First, the healthcare system has a moral and legal obligation to provide equal care to all patients, regardless of their race, ethnicity, or other characteristics. Second, discrimination in healthcare settings may cultivate patient disengagement from the healthcare system, thereby negatively affecting future healthcare encounters and patient health.[12,17,18,22,23,25,26] Finally, discrimination that is perceived within healthcare settings can potentially be addressed through quality improvement efforts implemented by healthcare systems, whereas there may be little healthcare systems can do in response to discrimination that occurs outside of their institutions.

For these reasons, it is important to determine the prevalence of perceived discrimination in healthcare settings and to understand its potential impact. Unfortunately, work in this area is inhibited by a wide variation in how perceived discrimination in healthcare settings has been measured across studies.[28] Most studies have measured

healthcare discrimination with either a single item from a scale of perceived discrimination[26] or with unique items developed for inclusion in a given survey.[6,7,12,20,23–25] Others have adapted existing, multi-item measures of perceived discrimination to assess discrimination that is perceived particularly within healthcare settings.[16,29] Still others have asked patients about their general perceptions of racism in health care rather than whether patients have personally experienced such discrimination.[13,17] The variation in measurement has made it difficult to draw conclusions about the prevalence and impact of discrimination in health care.

The current study was undertaken as an initial effort to compare the prevalence of perceived discrimination in health care across multiple measures within a single patient population. In a sample of 100 African American and White adults with diabetes, this exploratory study measured perceived discrimination using 3 types of measures that have often been used in prior research: a single-item assessing personal experiences with discrimination in health care (ie, personal discrimination, single-item),[30] a multiple-item measure adapted from a commonly-used measure of personal experiences with discrimination (ie, personal discrimination, multi-item),[29] and a measure assessing perceptions of general racism in the healthcare system, regardless of one's personal experiences with discrimination (ie, general racism).[17]

The primary aim was to examine differences in the prevalence of perceived discrimination in health care across patient race and type of measure. Based on previous findings that discrimination is more commonly experienced by racial and ethnic minorities than by Whites,[7,30] we predicted that African Americans would report more perceived discrimination than Whites across all measures of discrimination. Given that rates of perceived discrimination have been somewhat lower when assessed with a single-item than with multiple-items[7,29], we also predicted that the prevalence of personal discrimination for both Whites and African Americans would be higher on the multi-item than single-item measure. Finally, it has been well-established in the psychological literature that people are more likely to perceive discrimination against their group in general rather than against themselves personally.[31] We therefore predicted that rates of perceived discrimination would be higher on the general racism measure than on either the single-item or multi-item measure of personal discrimination.

A secondary aim was to explore whether each measure of perceived discrimination was associated with patient experiences with the healthcare system, including patients' perceptions of problems with their medical care and their receipt of screenings recommended for optimal diabetes management. This aim was included to explore whether conflicting evidence in the literature regarding whether perceived discrimination is associated with less patient satisfaction and/or healthcare utilization could be due to different measures of perceived discrimination being used across studies.[17,18,24,25]

Table 1. Sociodemographic characteristics of the sample

	African American (*n*=50)	White (*n*=50)	*P*-value*
Age, mean (range, SD)	63 (40–86, 11)	70 (52–85, 9)	.002
Income			.049
<$20K	50%	32%	
>$20K	40%	64%	
Missing	10%	4%	
High school, GED or less	54%	46%	.424

* t-test with unequal variances for age, chi-square test for income and education.

METHODS

Study Sample

The sample included 100 patients from the Veterans Affairs Pittsburgh Healthcare System (VAPHS) who were aged ≥18 years, self-identified their race as White or African American, had a diagnosis of diabetes from at least 2 years prior to the start of the study, and had no diagnosis of Alzheimer's or dementia (Table 1). To recruit the sample, a random sample of 479 patients (234 African Americans, 245 Whites) who met the inclusion criteria were identified from a VAPHS administrative database. Patients were mailed an initial letter and up to 2 follow-up letters inviting them to participate in a 30-minute telephone survey regarding their experiences with seeking treatment for diabetes. Consistent with local Institutional Review Board policies, only patients who indicated their interest in the study by mail or telephone were able to be enrolled in the study. African American (n=82, 35%) and White (n=93, 38%) patients were equally likely to express interest in the study ($\chi^2(1)=0.44$, $P=.51$). Trained research staff telephoned interested patients to explain the study in more detail, obtain patients' verbal informed consent, and administer the survey. Due to funding limitations, recruitment efforts ended after the goal of enrolling 100 patients had been met. Patients were compensated $20 for their participation.

Measures of Perceived Discrimination in Health Care

Personal Discrimination, Single-Item

An item from the validated and reliable Experiences of Discrimination

Table 2. Race differences in prevalence of perceived racial discrimination in health care across three types of measures

	African Americans (*n*=50)	Whites (*n*=50)	Unadjusted *P*-value†	Adjusted *P*-value‡
Personal discrimination, single-item measure (PD-S): While getting medical care, have you ever experienced discrimination, been prevented from doing something, or been hassled or made to feel inferior because of your race ethnicity, or color (% reporting that it occurred at least once)*	42	6	<.001	.004
Personal discrimination, multi-item measure (PD-M): When getting health care, how often has each experience happened to you because of your race or color (% reporting that experience occurred at least once)				
Treated with less courtesy than other people	30	4	.001	.032
Treated with less respect than other people	30	2	<.001	.027
Received poorer services than other people	30	6	.002	.026
Had a doctor or nurse act as if he or she thinks you were not smart	28	6	.004	.135
Had a doctor or nurse act as if he or she was afraid of you	22	2	.002	.151
Had a doctor or nurse act as if he or she was better than you	24	14	.202	.870
Felt like a doctor or nurse was not listening to what you were saying	34	16	.038	.498
% reporting any of the above*	42	18	.009	.134
General racism measure				
Doctors treat African American and White people the same. (% disagree or strongly disagree)	65	37	.007	.041
Racial discrimination in a doctor's office is common. (% agree or strongly agree)	38	14	.009	.032
In most hospitals, African American and Whites receive the same kind of care. (% disagree or strongly disagree)	66	25	<.001	.001
African Americans can receive the care they want as equally as White people can. (% disagree or strongly disagree)	53	33	.050	.270
% perceived discrimination on any of the above*§	74	40	.001	.004

* Variable used in regression models.
† Unadjusted *P* values comparing African American and white responses using chi-square tests.
‡ *P*-values for association of race with each measure, adjusting for patient age, education, and income, using logistic regression models.
§ Responses consistent with perceived discrimination were disagree or strongly disagree for items 1, 3, and 4, and agree or strongly agree for item 2.

(EOD) measure[30] was used as the single-item measure of personal discrimination in healthcare settings (Table 2). The original EOD asks how often (never, once, 2 or 3 times, 4 or more times) patients have encountered discrimination in 9 different settings, one of which is while getting medical care. The single medical care item has been used in previous studies to examine perceptions of discrimination in health care.[25,26] Because the current study focused on a VA patient population, the item was modified slightly so that it assessed patients' experiences of discrimination while getting medical care in either Veterans Affairs or non-Veterans Affairs facilities. Responses were dichotomized into never vs ever for analyses.

Personal Discrimination, Multi-Item

The multi-item measure of personal discrimination in healthcare settings was an adaptation of Williams' validated and widely-used Everyday Discrimination measure.[32,33] Williams' original measure assesses how often (never, once, 2 or 3 times, or 4 times or more) one has encountered 7 types of unfair treatment and the reason for the treatment (eg, race, sex). In previous studies, an adapted version was created specifically to assess race-based unfair treatment encountered within healthcare settings.[16,29] The healthcare-specific adapted version has shown excellent reliability in a variety of diverse patient populations[16,29] and was used in the current study (Table 2). We dichotomized each response into never vs ever and counted the number of items on which patients reported perceiving discrimination (coefficient alpha = .94). Preliminary analyses indicated that this count variable was not normally-distributed and that responses were best categorized into 2 levels (none vs any) for analyses.

General Racism in the Healthcare System

Perceptions that racism against African Americans exists in the healthcare system were measured by the 4-item Racism in Health Care Index (Table 2).[17] This was referred to as the ''general racism'' measure in the current study to highlight that it assesses perceptions of racism in health care regardless of patients' personal experi-

ences with such discrimination, which are assessed by the single-item and multi-item measures of personal discrimination. For this measure, patients were asked to indicate the extent to which they agree with four statements about racial discrimination in healthcare settings (strongly disagree, disagree, neither disagree or agree, agree, strongly agree). The number of statements on which patients perceived discrimination was calculated (coefficient alpha = .87). Preliminary analyses indicated that this count variable was not normally-distributed and that responses were best categorized into 2 levels (perceived discrimination on 0 vs >0 items) for analyses.

Sociodemographic Variables

Self-reported race, age, highest level of education completed, and income were assessed during the telephone survey.

Exploratory Outcome Measures

Perceptions of Problems with Health Care

Patients' perceptions of the care they receive for diabetes were measured using the 4-item Doctor-Patient Relationship subscale of the Questionnaire on Stress in Patients with Diabetes – Revised.[34] This scale assesses whether each of the following is a problem for the patient: different doctors give you different information regarding your diabetes; you feel insufficiently informed about your diabetes; doctors do not spend enough time with you; and your doctor does not treat your diabetes in the best possible way (coefficient alpha = .80). For analyses, we categorized patients into those who reported no problems vs at least one problem.

Receipt of Screenings for Diabetes Complications

Electronic medical records were examined to assess whether patients had received all recommended screenings for diabetes complications in the past 2 years. Based on standards from the Diabetes Quality Improvement Project, these included at least one hemoglobin A1c test, dilated eye exam, comprehensive foot exam, and urine protein test within the past year, and a fasting lipid test within the past 2 years.[35] Because all 5 tests are recommended for optimal diabetes management, patients were categorized as having received all 5 or fewer than 5 tests. Given that this outcome was based on Veterans Affairs medical records, patients who reported receiving no care from Veterans Affairs facilities in the past 12 months (n=9) were excluded from analyses of this outcome.

Statistical Analyses

Individual items and dichotomized summary scores of perceived discrimination measures were compared between races using chi-square tests for bivariate comparisons and logistic regression for tests of racial differences controlling for patient age, education, and income. Correlations among the dichotomized measures of perceived discrimination were examined in the total sample and within each race using phi coefficients. Mixed effect logistic regression models, which take into account the dependence of multiple outcomes within the same individual, were used to compare the dichotomized measures of perceived discrimination, adjusting for patient race, age, education, and income. Interactions between race and each perceived discrimination measure were tested and none were found to be significant, so models without these interactions are reported.

Separate regression models were used to test the association of each discrimination measure with patients' perceptions of care and receipt of recommended screenings for diabetes management. First, base models that included race, age, education, and income as predictors of each outcome were tested. The effects of perceived discrimination were then tested by adding each of the three measures one at a time to the base models. A criterion of $P<.05$ was used to determine statistical significance. Analyses were conducted using STATA/MP 10.1 (College Station, TX, 2008).

RESULTS

Sample Characteristics

The sample included 50 African Americans and 50 Whites, 99% of whom were male. Compared to Whites, African Americans were significantly younger and had lower incomes, but did not differ in educational attainment (Table 1).

Race Differences in Perceived Discrimination across Measures

Racial differences in the percentage of patients reporting discrimination on each measure of perceived discrimination in health care are reported in Table 2. As expected, African Americans were more likely than Whites to perceive discrimination in health care, although exact rates varied across individual items and dichotomized summary measures. Based on the personal discrimination, single-item measure, 42% of African Americans and 6% of Whites had experienced discrimination while getting medical care, a difference that was significant even after adjusting for patient age, education, and income (P=.004).

On individual items within the personal discrimination, multi-item measure, rates of perceived discrimination tended to be higher among African Americans (22%–34% across items) than among Whites (2%–16% across items). Although more African Americans than Whites perceived discrimination on at least one of the items, this difference was not statistically significant after controlling for patient characteristics (42% vs 18%, P=.134).

Table 3. Phi correlation coefficients among perceived discrimination measures for total sample and each racial group

Measures of perceived discrimination	Total Sample		African Americans		Whites	
	PD-S	PD-M	PD-S	PD-M	PD-S	PD-M
PD-M	.55†		.51†		.54†	
General racism	.49†	.48†	.50†	.50†	.31*	.36*

PD-S: personal discrimination, single-item measure.
PD-M: personal discrimination, multi-item measure.
* $P<.05$.
† $P<.001$.

Finally, on most of the individual items assessing general racism in the healthcare system, rates of perceived discrimination were higher among African Americans (38%–66% across items) than among Whites (14%–37% across items). African Americans were significantly more likely than Whites to perceive discrimination on at least one of the 4 individual items (74% vs 40%, respectively, $P=.004$).

Comparisons across Measures of Perceived Discrimination

As shown in Table 3, the three dichotomized measures of perceived discrimination were significantly positively correlated with one another in the total sample and within each racial group. When all three measures were included in a mixed effect regression model (# of observations = 279, # of patients = 93), there was a significant effect for race such that African Americans reported more perceived discrimination than Whites (OR= 18.57, 95% CI= 2.43–142.04). Rates of discrimination also differed across measures, such that patients were more likely to report perceived discrimination on the general racism measure compared to the personal discrimination, single-item (OR=36.53, 95% CI=7.95–167.89) or multi-item (OR=20.28, 95% CI=5.12–80.34) measures. Responses were not significantly different on the personal discrimination, single-item and personal discrimination, multi-item measures (OR=1.80, 95% CI=.61–5.32).

Perceived Discrimination and Patients' Perceptions of Care

Overall, 56% of patients reported at least one problem with their diabetes care (54% and 58% of African Americans and Whites, respectively). In a model containing only patient characteristics as predictors, the likelihood of reporting a problem was not significantly associated with patient race, education, or income, but did decline with age (OR=.95, 95% CI=.91–1.00, $P=.047$; Table 4, Model 0). Models in which each measure of perceived discrimination was separately added to the base model indicated that the likelihood of patient-reported problems was not significantly associated with the personal discrimination, single-item measure (OR=2.57, 95% CI=.77–8.56; Table 4, Model 1). However, the personal discrimination, multi-item (OR=3.96, 95% CI=1.29–12.18; Table 4, Model 2) and general racism (OR=3.61, 95% CI=1.34–9.71; Table 4, Model 3) measures were each associated with nearly a four-fold increase in the likelihood of reporting a problem with care.

Perceived Discrimination and Patients' Receipt of Recommended Screenings

Patients received a median of 4 (IQR = 3–5) out of 5 recommended

Table 4. Logistic regression models testing association of perceived discrimination measures with patient-reported problems with diabetes care, adjusting for sociodemographic characteristics

Predictors	Model 0		Model 1		Model 2		Model 3	
	OR	(95% CI)	OR	(95% CI)	OR	(95% CI)	OR	(95% CI)
Sociodemographics								
African American race	0.46	(0.17–1.22)	0.35	(0.12–1.01)†	0.36	(0.13–1.02)†	0.29	(0.10–0.85)*
Age	0.95	(0.91–1.00)*	0.96	(0.91–1.00)†	0.96	(0.91–1.01)†	0.95	(0.91–1.00)*
Education (>high school)	1.28	(0.51–3.24)	1.38	(0.54–3.56)	1.20	(0.46–3.11)	1.30	(0.50–3.38)
Income <$20,000	0.41	(0.16–1.04)†	0.43	(0.17–1.11)†	0.49	(0.19–1.29)	0.43	(0.17–1.14)†
Perceived discrimination								
PD-S			2.57	(0.77–8.56)				
PD-M					3.96	(1.29–12.18)*		
General racism							3.61	(1.34–9.71)*
Pseudo R^2	.07†		.09*		.12*		.12*	

$N=93$ for all models.
PD-S: personal discrimination, single-item measure.
PD-M: personal discrimination, multi-item measure.
* $P<.05$.
† $P<.10$.

screenings for diabetes complications within the prior two years, with 35% of both Whites and African Americans having received all 5 tests. Receiving all 5 tests was not significantly associated with patient characteristics or with any of the measures of perceived discrimination (data not shown).

DISCUSSION

This study compared rates of perceived racial discrimination in health care among African American and White patients using three different measures of discrimination. The measures were moderately correlated with one another, suggesting that they tap a similar underlying construct. However, rates of perceived discrimination were higher when assessed using a measure of general racism in health care than when assessed using a single-item or multi-item measure of personal experiences with discrimination in healthcare settings. This is consistent with the robust finding in psychological literature that people report less discrimination directed at themselves than at their group in general.[31] The current study indicates that this personal/group discrimination discrepancy extends to healthcare discrimination.

Rates of personal discrimination in healthcare settings did not differ significantly when assessed using a single-item or multi-item measure. This is somewhat surprising, given that rates of perceived discrimination in health care tend to be lower in studies using only a single item to assess discrimination than in those using multi-item measures.[7,22,25,26,29] Studies using single-item measures have found that 6%–12% of African American patients report having experienced racial discrimination in healthcare settings,[7,12,13,22,25,26] whereas discrimination was reported by 63% of African American patients in a study using a multi-item measure.[29] Past studies, however, have not compared single-item and multi-item measures within the same patient population, as in the current study.

A secondary goal of the current study was to explore whether the measure used to assess discrimination in health care influences the degree to which perceived discrimination is associated with patients' experiences with the healthcare system. It has been proposed that experiences of discrimination foster disengagement from the healthcare system.[25] Several studies have reported significant associations between perceived discrimination and indicators of patient disengagement (eg, delaying necessary care),[12,15–18,20,23] but conflicting evidence has also been reported.[12,16,18,24–26] Most studies that have found no association,[24–26] or inconsistent relationships across outcomes,[12,18,23] have relied on single-item measures of perceived discrimination, which may have contributed to the inconsistent findings. In the current study, a multi-item measure assessing personal discrimination encountered in healthcare settings was associated with patient-reported problems with their care, whereas a single-item measure failed to predict this outcome.

None of the measures in the current study were associated with patients' receipt of screenings recommended for optimal diabetes management, even though previous studies have reported that perceived discrimination among patients with diabetes is associated with a lower likelihood of obtaining A1c tests, eye exams, and diabetic foot exams.[12,23] Past studies relied on patient self-report to assess screening behavior whereas the current study obtained this information from medical records.

Several limitations should be considered when interpreting the findings of this study. This study was designed as an exploratory study with a small sample and, as such, is not powered to detect small effects or interactions among the variables. The sample size, coupled with the low prevalence of perceived discrimination reported by Whites, required the summary measures of discrimination to be dichotomized, which may have further limited the study's statistical power. The nature of the study sample, which consisted of older, primarily male Veterans with diabetes who were recruited by mail from a single Veterans Affairs facility, also constrained the generalizability of our findings. The differences found across measures in this sample, however, suggest the importance of examining these issues in a larger study with a more representative group of patients.

Although the study compared the 3 types of measures most commonly used in research on perceived discrimination in health care, it was not possible to assess the impact of every factor that could affect reports of discrimination, such as the timeframe in which patients experienced discrimination (eg, in the past 12 months vs ever) or whether the instruments are self-administered or interviewer-administered.[1] Furthermore, the measures of discrimination in this study assessed patients' perceptions of discrimination rather than verifiable instances of discrimination. It is therefore unclear whether the rates of discrimination reported by patients are over-, under-, or accurate estimates of patients' actual encounters with discrimination. However, patient perceptions of discrimination are likely to influence their reactions or behavior in a given situation,[36] regardless of whether discrimination has objectively occurred, thereby justifying the focus on perceived rather than actual discrimination.

This study also focused on only two measures of patient experiences with the healthcare system, including patient-reported problems with their care and screening behavior. There may be other aspects of patient experiences with the healthcare system that are more sensitive to patients' perceptions of discrimination that were not examined in this study. The study's cross-sectional design

also does not allow one to draw causal conclusions about the relationships observed between measures of perceived discrimination and patients' receipt of care.

These limitations notwithstanding, this study makes a notable contribution to existing literature on perceived discrimination and health care by examining how the prevalence of perceived discrimination in healthcare settings and its association with patient experiences depends on how perceived discrimination is measured. The current study suggests that measures that assess patients' perceptions of general racism in the healthcare system yield considerably higher rates of discrimination than measures that assess patients' personal experiences with such discrimination. Moreover, single-item measures of patients' personal experiences with discrimination in healthcare settings are less likely to be associated with patient-reported problems with their medical care than a multi-item measure of personal discrimination.

Based on these findings, healthcare systems should consider assessing patients' experiences with discrimination and how they relate to outcomes of interest using the multi-item measure of personal discrimination from the current study. Collecting this information as part of standard quality control activities would allow healthcare systems to monitor the extent to which their patients perceive discrimination while obtaining services. Assessing whether perceived discrimination is a problem for a given healthcare system is an important first step towards developing strategies to address it. Regularly assessing patients' perceptions of healthcare discrimination using a reliable, sensitive measure could provide valuable information to guide patient outreach or provider education activities designed, in part, to reduce perceived discrimination in health care and its negative influence on patients' perceptions of care.

ACKNOWLEDGMENTS

This material is the result of work supported by resources and the use of facilities at the Veterans Affairs Pittsburgh Healthcare System in Pittsburgh, PA. The authors take full responsibility for the contents of this paper, which do not represent the views of the Department of Veterans Affairs or the United States Government. The study was funded by the Center for Health Equity Research and Promotion Pilot Research Program (LIP 72-035). Dr. Hausmann was also supported by VA Health Services Research and Development Career Development Awards (RCD 06-287 & ER 0280-1). Dr. Ibrahim is the recipient of a VA Health Services Research Career Development Award and the Harold Amos Robert Wood Johnson Scholar Award. Dr. Ibrahim was also supported by a K24 Award (1K24AR055259-01) from the National Institutes of Musculoskeletal and Skin Disorders. Dr. Kressin was supported by a Research Career Scientist award, Department of Veterans Affairs, Health Services Research & Development (RCS 02-066-1). The authors thank Amie Cande and Elizabeth Flatley for their assistance with data collection. The authors have no conflicts of interest to disclose.

REFERENCES

1. Paradies Y. A systematic review of empirical research on self-reported racism and health. *Int J Epidemiol*. 2006;35(4):888–901.
2. Williams DR, Mohammed SA. Discrimination and racial disparities in health: evidence and needed research. *J Behav Med*. 2009;32(1):20–47.
3. Pascoe EA, Smart Richman L. Perceived discrimination and health: a meta-analytic review. *Psychol Bull*. 2009;135(4):531–554.
4. Smedley BD, Stith AY, Nelson AR, eds. *Unequal treatment: Confronting racial and ethnic disparities in health care*. Washington, DC: National Academies Press; 2003.
5. Krieger N. Embodying inequality: a review of concepts, measures, and methods for studying health consequences of discrimination. *Int J Health Serv*. 1999;29(2):295–352.
6. Wagner J, Abbott G. Depression and depression care in diabetes: relationship to perceived discrimination in African Americans. *Diabetes Care*. 2007;30(2):364–366.
7. Hausmann LR, Jeong K, Bost JE, Ibrahim SA. Perceived discrimination in health care and health status in a racially diverse sample. *Med Care*. 2008;46(9):905–914.
8. Barnes LL, de Leon CF, Lewis TT, Bienias JL, Wilson RS, Evans DA. Perceived discrimination and mortality in a population-based study of older adults. *Am J Public Health*. 2008;98(7):1241–1247.
9. Borrell LN, Jacobs DRJr, Williams DR, Pletcher MJ, Houston TK, Kiefe CI. Self-reported racial discrimination and substance use in the Coronary Artery Risk Development in Adults Study. *Am J Epidemiol*. 2007;166:1068–1079.
10. Schulz AJ, Gravlee CC, Williams DR, Israel BA, Mentz G, Rowe Z. Discrimination, symptoms of depression, and self-rated health among African American women in Detroit: results from a longitudinal analysis. *Am J Public Health*. 2006;96(7):1265–1270.
11. Ryan AM, Gee GC, Laflamme DF. The association between self-reported discrimination, physical health and blood pressure: findings from African Americans, Black immigrants, and Latino immigrants in New Hampshire. *J Health Care Poor Underserved*. 2006;17(2 Suppl):116–132.
12. Trivedi AN, Ayanian JZ. Perceived discrimination and use of preventive health services. *J Gen Intern Med*. 2006;21(6):553–558.
13. Lillie-Blanton M, Brodie M, Rowland D, Altman D, McIntosh M. Race, ethnicity, and the health care system: public perceptions and experiences. *Med Care Res Rev*. 2000;57 Suppl 1:218–235.
14. LaVeist TA, Rolley NC, Diala C. Prevalence and patterns of discrimination among U.S. health care consumers. *Int J Health Serv*. 2003;33(2):331–344.
15. Van Houtven CH, Voils CI, Oddone EZ, et al. Perceived discrimination and reported delay of pharmacy prescriptions and medical tests. *J Gen Intern Med*. 2005;20:578–583.
16. Bird ST, Bogart LM, Delahanty DL. Health-related correlates of perceived discrimination in HIV care. *AIDS Patient Care and STDs*. 2004;18(1):19–26.
17. LaVeist TA, Nickerson KJ, Bowie JV. Attitudes about racism, medical mistrust, and satisfaction with care among African American and White cardiac patients. *Med Care Res Rev*. 2000;57(Suppl 1):146–161.
18. Blanchard J, Lurie N. R-E-S-P-E-C-T: patient reports of disrespect in the health care setting and its impact on care. *J Fam Pract*. Sep 2004;53(9):721–730.
19. Blendon RJ, Buhr T, Cassidy EF, et al. Disparities in health: perspectives of a multi-ethnic, multi-racial America. *Health Aff (Millwood)*. 2007;26(5):1437–1447.
20. Wamala S, Merlo J, Bostrom G, Hogstedt C. Perceived discrimination, socioeconomic disadvantage and refraining from seeking medical treatment in Sweden. *J Epidemiol Community Health*. 2007;61(5):409–415.

21. Gee GC, Ryan A, Laflamme DJ, Holt J. Self-reported discrimination and mental health status among African descendants, Mexican Americans, and other Latinos in the New Hampshire REACH 2010 Initiative: the added dimension of immigration. *Am J Public Health*. 2006;96(10):1821–1828.
22. Piette JD, Bibbins-Domingo K, Schillinger D. Health care discrimination, processes of care, and diabetes patients' health status. *Patient Educ Couns*. 2006;60(1):41–48.
23. Ryan AM, Gee GC, Griffith D. The effects of perceived discrimination on diabetes management. *J Health Care Poor Underserved*. 2008;19(1):149–163.
24. Hausmann LR, Jeong K, Bost JE, Ibrahim SA. Perceived discrimination in health care and use of preventive health services. *J Gen Intern Med*. 2008;23(10):1679–1684. PMCID: PMC2533365.
25. Burgess DJ, Ding Y, Hargreaves M, van Ryn M, Phelan S. The association between perceived discrimination and underutilization of needed medical and mental health care in a multi-ethnic community sample. *J Health Care Poor Underserved*. 2008;19(3):894–911.
26. Casagrande SS, Gary TL, Laveist TA, Gaskin DJ, Cooper LA. Perceived discrimination and adherence to medical care in a racially integrated community. *J Gen Intern Med*. 2007;22(3):389–395.
27. King RK, Green AR, Tan-McGrory A, Donahue EJ, Kimbrough-Sugick J, Betancourt JR. A plan for action: key perspectives from the racial/ethnic disparities strategy forum. *Milbank Q*. 2008;86(2):241–272.
28. Kressin NR, Raymond KL, Manze M. Perceptions of race/ethnicity-based discrimination: a review of measures and evaluation of their usefulness for the health care setting. *J Health Care Poor Underserved*. 2008;19 (3):697–730.
29. Bird ST, Bogart LM. Perceived race-based and socioeconomic status (SES)-based discrimination in interactions with health care providers. *Ethn Dis*. 2001;11(3):554–563.
30. Krieger N, Smith K, Naishadham D, Hartman C, Barbeau EM. Experiences of discrimination: validity and reliability of a self-report measure for population health research on racism and health. *Soc Sci Med*. 2005;61(7): 1576–1596.
31. Taylor DM, Wright SC, Porter LE. Dimensions of perceived discrimination: The personal/group discrimination discrepancy. In: Zanna MP, Olson JM, eds. *The Psychology of Prejudice: The Ontario Symposium*. Vol 7. Hillsdale, NJ: Erlbaum; 1994:233–255.
32. Williams DR, Yu Y, Jackson JS, Anderson NB. Racial differences in physical and mental health: socio-economic status, stress and discrimination. *J Health Psychol*. 1997;2:335–351.
33. Taylor TR, Kamarck TW, Shiffman S. Validation of the Detroit Area Study Discrimination Scale in a community sample of older African American adults: the Pittsburgh healthy heart project. *Int J Behav Med*. 2004;11(2):88–94.
34. Herschbach P, Duran G, Waadt S, Zettler A, Amm C, Marten-Mittag B. Psychometric properties of the Questionnaire on Stress in Patients with Diabetes–Revised (QSD-R). *Health Psychol*. 1997;16(2):171–174.
35. McLaughlin S. Diabetes Quality Improvement Project Initial Measure Set (Final Version). Available at: http://journal.diabetes.org/diabetesspectrum/00v13n1/pg5. Last accessed: October 19, 2009.
36. Clark R, Anderson NB, Clark VR, Williams DR. Racism as a stressor for African Americans: A biopsychosocial model. *American Psychologist*. 1999;54(10):805–816.

AUTHOR CONTRIBUTIONS

Design concept of study: Hausmann, Kressin, Ibrahim
Acquisition of data: Hausmann
Data analysis and interpretation: Hausmann, Kressin, Hanusa
Manuscript draft: Hausmann, Kressin, Hanusa, Ibrahim
Statistical expertise: Hausmann, Hanusa
Acquisition of funding: Hausmann, Ibrahim
Administrative, technical, or material assistance: Hausmann, Hanusa, Ibrahim
Supervision: Kressin, Ibrahim

Addiction

RESEARCH REPORT

doi:10.1111/j.1360-0443.2009.02880.x

The effects of binge drinking on college students' next-day academic test-taking performance and mood state

Jonathan Howland[1], Damaris J. Rohsenow[2], Jacey A. Greece[3], Caroline A. Littlefield[1], Alissa Almeida[4], Timothy Heeren[1], Michael Winter[1], Caleb A. Bliss[1], Sarah Hunt[1] & John Hermos[1]

Youth Alcohol Prevention Center at Boston University School of Public Health, Boston, MA, USA[1] Center for Alcohol and Addiction Studies at Brown University, Providence, RI, USA,[2] Abt Associates, Cambridge, MA, USA[3] and Boston Medical Center, Boston, MA, USA[4]

ABSTRACT

Aim To assess the effects of binge drinking on students' next-day academic test-taking performance. **Design** A placebo-controlled cross-over design with randomly assigned order of conditions. Participants were randomized to either alcoholic beverage [mean = 0.12 g% breath alcohol concentration (BrAC)] or placebo on the first night and then received the other beverage a week later. The next day, participants were assessed on test-taking, neurocognitive performance and mood state. **Participants** A total of 196 college students (≥21 years) recruited from greater Boston. **Setting** The trial was conducted at the General Clinical Research Center at the Boston Medical Center. **Measurements** The Graduate Record Examinations© (GREs) and a quiz on a lecture presented the previous day measured test-taking performance; the Neurobehavioral Evaluation System (NES3) and the Psychomotor Vigilance Test (PVT) measured neurocognitive performance; and the Profile of Mood States (POMS) measured mood. **Findings** Test-taking performance was not affected on the morning after alcohol administration, but mood state and attention/reaction-time were affected. **Conclusion** Drinking to a level of 0.12 g% BrAC does not affect next-day test-taking performance, but does affect some neurocognitive measures and mood state.

Keywords Academic performance, binge drinking, intoxication, mood state, neurocognitive performance, students.

Correspondence to: Jonathan Howland, Department of Community Health Sciences, Boston University School of Public Health, 801 Massachusetts Avenue, Crosstown Center—3rd Floor, Boston, MA 02118, USA. E-mail: jhowl@bu.edu

Submitted 27 July 2009; initial review completed 30 September 2009; final version accepted 29 October 2009

INTRODUCTION

The National Advisory Council of the National Institute on Alcohol Abuse and Alcoholism (NIAAA) defines binge drinking as attaining a blood alcohol concentration (BAC) of 0.08 g% or more, corresponding, for most adults, to five or more drinks (more than four if female) in about 2 hours [1]. In the United States, both binge drinking and heavy drinking (binge drinking at least five times in the last 30 days [1]) peak at age 21 [2].

Although college students have lower rates of daily drinking than their non-college peers, they have higher rates of binge drinking [3], with 32–44% reporting binge drinking [4]. Not surprisingly, 60–75% of college students experience at least one hangover a year, 27% report one to two hangovers and 34% report 12–51 hangovers [5].

Serious negative consequences associated with student drinking include death [6], injury, suicide, fighting, unprotected sex, rape, property damage, and legal problems; academic difficulties are, however, the most frequently reported consequence of excessive student drinking [7]. Academic problems resulting from heavy drinking can occur through several mechanisms: hangover results in missing morning classes; drinking uses time otherwise spent studying; drinking impedes next-day learning in class or, when studying, by affecting memory retention [8]; and personal and interpersonal problems resulting from heavy drinking may make it difficult to focus on school work [9,10].

A number of surveys have shown relationships between college students' drinking and academic difficulties [7,9–15]. Other survey studies, however, have found that the relationship of drinking and academic performance disappeared after controlling for pre-college differences in academic performance [16,17].

Little experimental work has been published on the effects of student drinking on academic performance. There is, however, a body of experimental research on the effects of intoxication on next-day performance ('residual effects of alcohol'), as measured by neurocognitive laboratory tests or occupational training simulators. Because academic performance is the occupation of students, this research is relevant to the question of whether intoxication in the evening impairs students' next-day test-taking ability, when blood alcohol concentration (BAC) has returned to zero. Several studies found residual alcohol effects on simulated occupational tasks [18–29]. However, in other experimental studies residual effects of intoxication were not found for occupational tasks [30–34]. Some investigators have found residual alcohol effects on various neurocognitive tests [35–44], but other studies found no impairment on tests of manual dexterity or neurocognitive performance [39,45–49].

Inconsistencies among study findings may be the result of factors such as the type of performance measured the amount of alcohol administered, the age and alcohol tolerance of participants and the length of time from drinking to testing [49].

We conducted a randomized cross-over trial to examine the extent to which alcohol intoxication affects college students' next-day academic performance at zero BAC. Neurocognitive tasks relevant to academic performance were also assessed. We hypothesized that drinking to about 0.12 g% BrAC would not affect next-day performance on academic tests requiring long-term memory (e.g. standardized academic achievement tests), but would affect performance on tests of recently learned material and on neurocognitive tests requiring sustained attention and speed. To our knowledge, this is the first study to explore experimentally the relationship between binge drinking and academic performance.

METHODS

Participants

Participants were university students recruited from greater Boston, Massachusetts, who were between 21 and 24 years of age and met the following criteria: (1) no drinking problems (score <5 on the Short Michigan Alcohol Screening Test (SMAST)) [50] and no history of treatment or counseling for chronic alcohol problems; (2) consumption of more than five drinks (more than four if female) on a single occasion at least once in the 30 days prior to screening; (3) no health problems or current medication use contraindicated for alcohol; (4) no diagnosis of sleep disorders or use of sleeping medications; (5) fluent English; (6) recently graduated from, or currently attending, an institution of higher learning; (7) not working night shifts; (8) not a daily smoker; (9) not traveled across two or more time zones in the prior month; and (10) if female, negative pregnancy test and not nursing. Female participants' menstrual cycle phase was documented, but not a factor in scheduling their experimental sessions [51–53]. For safety reasons, regular tobacco users were excluded because participants were not allowed to leave the laboratory to smoke. This exclusion also avoided possible confounding due to nicotine withdrawal during the study sessions. Before beverage administration, participants who reported consuming alcohol, caffeine, prescription or over-the-counter drugs within the prior 24 hours, or who had had a positive breath alcohol test (BrAC), were rescheduled (see Table 1 for participant characteristics).

No information about individuals' participation was provided to institutions attended by volunteers. Participants were paid $300 upon completion of the study, or a pro rata amount if their participation ended prior to completing the study. The Institutional Review Boards at Boston Medical Center and Brown University approved this study.

Study design

We used a placebo-controlled, double-blind, within-subjects, repeated-measures design to study the residual effects of alcohol, with participants serving as their own controls. Participants took part in the study over 4 days: an evening and the next morning, followed a week later by the same schedule. All participants received two beverages (alcohol and placebo) in counterbalanced order (alcohol week 1 versus alcohol week 2).

Study procedures

Recruitment and screening

Participants were recruited by advertisements in local newspapers and websites (e.g. Facebook and Craig's List). Interested individuals were first screened by telephone and then in person, including a physician examination (after informed consent). To reduce potential confounding by sleep pattern variations, participants were instructed to keep a sleep diary, comply with a minimum regimen of 8 hours sleep (retiring to bed no later than midnight and awaking no later than 8 a.m.), with confirmation call-ins to a time-stamped answering machine each evening and morning for the 3 nights prior to

Table 1 Participant characteristics.

Total (n = 193)	
Sex	
Male	107 (55.4%)
Female	86 (44.6%)
Age	
Mean ± SD	21.47 (0.64)
Range	21–24
Race	
White	155 (80.3%)
Black	8 (4.2%)
Asian	13 (6.7%)
Other	17 (8.8%)
Family history of alcohol problems	
Yes	71 (36.8%)
No	119 (61.7%)
Adopted	3 (1.6%)
Mean age of drinking onset	
Mean ± SD	16.18 (1.66)
Range	11–21
Maximum breath alcohol concentration (BrAC)	
Mean ± SD	0.12 (0.01)
Range	0.09–0.16
Amount of alcohol received (ml)	
Male: mean ± SD	1609 (288)
Male: range	1052–2308
Female: mean ± SD	1122 (178)
Female: range	683–1606
% with hangover	
Rated hangover >1 on the morning following alcohol administration when asked to rate their hangover on a scale from 0 (no hangover) to 7 (incapacitating hangover)	69.8%
Morning mean AHS score	
Placebo condition	0.71 (0.35)
Alcohol condition	1.38 (0.81)

AHS: Acute Hangover Scale; SD: standard deviation.

Table 2 Schedule of study procedures.

Orientation/consent 10 a.m.–12 p.m.	Orientation. Consent. Enrollment questionnaires. Medical screening by physician
Evening sessions 4 p.m.–5 p.m.	Dinner, screened for adherence to study protocol. BrAC tested. Pregnancy tests administered to females
5 p.m–6 p.m	Family Tree questionnaire administered. Practice tests to familiarize participants with GRE and PVT
6 p.m–7.30 p.m	Video lecture based on next-day's quiz. Participants study lecture notes for 1 hour
7.30 p.m–8.45 p.m	Practice NES3 test
8.45 p.m–11 p.m	Beverage administration
	Repeated BrAC tests
11 p.m	Lights out
	Observed throughout night by EMT
Morning sessions 7 a.m.–7.30 a.m.	Subjects awakened. Morning questionnaires
7.30 a.m.–8 a.m.	Breakfast
8 a.m.–11 a.m.	BrAC tests
	POMS questionnaire, quiz on video lecture, GRE, NES3, PVT, self-rated performance questionnaire
12.30 p.m.	Subjects dismissed

BrAC: breath alcohol concentration; GRE: Graduate Record Examinations; PVT: Psychomotor Vigilance Test; EMT: emergency medical technician; POMS: Profile of Mood States; NES3: Neurobehavioral Evaluation System.

experimental sessions. Participants were told not to nap and, for 24 hours prior to their experimental sessions, to abstain from alcohol, medications not already approved by the study physician, sleep aids, recreational drugs and caffeine. To familiarize participants with the standard academic achievement tests, they were required to read and complete a practice booklet issued by the testing service.

One week after screening and enrollment, participants returned in groups of three to five for the first overnight experimental session. They reported at 4 p.m.; car keys were collected from participants who drove to the study site; compliance with pre-laboratory regimens was checked; and, following a standardized dinner, participants were screened for zero breath alcohol (BrAC) and negative pregnancy test (if female). To prepare for a quiz the following morning, at 6 p.m. participants viewed randomly one of two 30-minute video lectures on a public health topic and had an hour to study an accompanying textbook chapter. They viewed the other video lecture the following week. To reduce potential learning effects, participants then practised the computer-based neurocognitive test prior to alcohol administration (Table 2).

Randomization procedures

For the first experimental session, participants received a study ID number and were assigned randomly to beverage (placebo or alcohol); they received the other beverage the following week. For safety reasons, no more than three of the five participants received alcohol on any given night. To maintain double blinding, the individual who prepared beverages and conducted breath tests had no other contact with participants; all other study assistants working directly with participants were unaware of participants' beverage assignments. Participants were told there was a 50–50 chance of receiving alcohol the first night and they were instructed not to inspect or taste each others' drinks or discuss the beverage they received.

Beverage administration procedures

Alcoholic beverage administration targeted 0.12 g% BrAC, adjusting the alcohol per kilogram of body weight for sex (1.068 g/kg body weight for men and 0.915 g/kg for women), as per Friel *et al.* [54]. Males received a mean of 1609.07 (SD: 288.55) ml of beverage (range: 1052.20–2308.00), or the equivalent of 6.75 12-oz cans of regular beer (at 4.82% alcohol by volume); females received a mean of 1122.09 (SD: 178.48) ml of beverage (range: 683.3–1606.60), or the equivalent of 4.72 12-oz cans of regular beer.

Beer controlled with non-alcoholic beer has been shown to be one of the two most effective beverage combinations for disguising placebo [55]. Beer was chosen because most young men and women find it palatable. Elephant Beer™ (Carlsberg, Copenhagen V, Denmark) with 7.2% alcohol and Clausthaler™ non-alcoholic beer (Radeberger Gruppe KG, Frankfurt am Main, Germany) were the beverages. High alcohol beer reduces the volume required to achieve the targeted BrAC. Beverage administration began 4 hours after eating and went from 8.45 p.m. to 9.45 p.m. (up to 10.00 p.m. as needed). Participants were told the total number of cups of beverage they were to consume in an hour. They were asked to drink the first two cups (330–340 ml) quickly and to pace the rest over the time allowed. Participants were breath tested 15 minutes after completing their beverage. If participants randomized to alcohol did not reach 0.12 g% BrAC, the ratio of obtained versus targeted BrAC was used to estimate the additional amount of beer to be administered. To maintain blinding, some of the placebo participants were given a matched extra dose of non-alcoholic beer. After participants finished drinking, they were breath tested every 15 minutes prior to bedtime, with the last BrAC measurement recorded 5 minutes before lights out.

Following beverage administration and a 30-minute absorption period, participants completed questionnaires, received snacks and prepared for bed. Participants had an 8-hour opportunity to sleep (no lights or television and cellphones turned off) between 11 p.m. and 7 a.m. in an individual bedroom with bathroom. They were monitored throughout the night for safety by an emergency medical technician (EMT).

At 7 a.m. participants were awakened, breath-tested and served breakfast (no caffeine). They then completed a questionnaire assessing mood state and, at 8 a.m., started testing. Sleep inertia during the first 30 minutes after waking is likely to impair performance [56]; allowing an hour before performance testing avoids this. To avoid confounding by alcohol remaining in the blood, performance testing was delayed, if necessary, until BrAC reached <0.00 g%. Participants were dismissed from this session at approximately 11.30 a.m. They were given an additional mood assessment questionnaire in a self-addressed, postage-paid return envelope and asked to complete it at 5 p.m. that day and mail it back to the study coordinator. One week later they returned for the second experimental session, identical except for beverage, video lecture and the standardized test version.

Individual difference measures

Recent drinking practice was estimated using a two-item alcohol use questionnaire: (i) 'Considering all your drinking times in the past 30 days, about how often did you have any beer, wine or liquor?', Likert-rated from 1 'once a day' to 7, 'did not drink', with each point anchored; and (ii) 'In the past 30 days, on a typical day that you drank, about how much did you have to drink in one day?', rated from 1 to 8, with choices of one to seven drinks and 'eight or more drinks'. (One drink was defined as 12 ounces of beer or wine cooler, 4 oz of wine or 1 oz of liquor.) Average daily volume (ADV) was calculated as the product of these. We also collected information on family history of drinking problems using the Family History Tree questionnaire developed by Mann *et al.* [57] and on age of drinking onset. These data are presented in Table 1, but were not included in th analyses.

Dependent measures of objective effects

Overview

Two tests of academic performance were used. Short-term recall was assessed by a quiz on a lecture delivered prior to beverage administration. Versions of the Graduate Record Examinations© (GREs) (Educational Testing Service, Princeton, NJ) were used to measure verbal and quantitative skills that have been acquired over a long period of time. Two methods of assessing neurocognitive performance were used: the Neurobehavioral Evaluation System (NES3), a neurocognitive battery; and the Psychomotor Vigilance Task (PVT), a measure of sustained attention/reaction-time.

Lecture quiz

First we administered a 30-question quiz based on the videotaped lecture and associated reading presented the day before. Two lectures and readings were used in counterbalanced order. The two lectures were based on chapters from a public health text, *Introduction to Public Health* [58]: Chapter 15, 'Tobacco: Public Health Threat Number One' and Chapter 16, 'Diet and Activity: Public Health Threat Number Two'. Quiz questions were derived from the teacher's guide. The quizzes were pilot-tested previously with 50 college students to ensure a normal distribution of scores.

GREs

After the quiz, we administered two parts of the GRE's General Test: a 30-minute verbal section (ability to discern, comprehend and analyze words, sentences and written passages) and a 45-minute quantitative section (basic mathematical skills, elementary mathematical concepts and ability to reason and to solve quantitative problems) in four broad content areas: arithmetic, algebra, geometry and data analysis [59]. Two different, but comparable, computer-administered and computer-scored tests were used, with order randomized by individual.

For assessments, participants had their own carrels and were monitored to ensure that they did not communicate. To enhance motivation, participants who scored in the top 50% of national averages on both sections received up to four complimentary movie tickets (two per study week). Participants were not informed of their scores or awarded tickets until they had completed the study.

NES3

The NES3 is a computer-assisted battery of cognitive tests validated for cognitive impairment [60]. As primary measures, we selected nine tests requiring speed, sustained attention or sustained attention/reaction-time, tests most apt to be affected the day after intoxication [61]. For manual dexterity tests that tested each hand individually, we used the test for the preferred hand; for tests that had forward and backward versions, we used the more difficult backward versions. The following tests assessed speed: Finger Tapping Test, preferred hand (FTT-P) (assesses manual motor speed and dexterity); and Sequences Test A, latency (ST-A-L); Digit-Symbol Test, latency (DST-L); Pattern Memory Test, latency (PMT-L) (all assessing speed of cognitive processing). The following tests assessed sustained attention: Auditory Digit Span Test, backwards (ADST-B); Adaptive Paced Auditory Serial Addition Test, number correct (APASAT-C); Visual Span Test, backward (VST-B); Pattern Memory Test, number correct (PMT-C). The Continuous Performance Test (CPT) measures both sustained attention and reaction-time.

PVT

As an additional test of sustained attention/reaction-time, we used the Psychomotor Vigilance Task [62] (Ambulatory Monitoring, Inc, Ardsley, NY, USA). On this hand-held unit participants press a button with their preferred hand as quickly as possible in response to numbers scrolling on an LCD screen, with a random 3–7-second interstimulus interval. Response time is counted in milliseconds. A solid-state storage unit collects data for downloading to a PC. The recorded outcome variable is median reaction-time.

Exploratory measures

As exploratory measures, we administered an additional nine NES-3 tests: FTT (non-preferred hand); ST (backward); ADST-F (forward); APASAT (stimulus response rate); VST (forward); VT (Vocabulary Test, a measure of general verbal ability); LOT (Line Orientation Test, number correct and latency, both measures of attention to visiospatial information); and LL (List Learning, a measure of quantitative aspects of several components of verbal learning and memory).

Dependent measures of subjective effects

Mood

Because the residual effects of alcohol on mood state might be salient to college students, we also measured next-day mood in both the morning and the afternoon. To assess mood, we used the Profile of Mood State Brief Form (POMS) [63], a validated self-administered questionnaire with 30 adjectives [each rated on a five-point Likert scale, from 0 (not at all) to 4 (extremely)]. These comprise six domains: fatigue–inertia (F); tension–anxiety (T); depression–dejection (D); anger–hostility (A); confusion–bewilderment (C); and vigor–activity (V). Only total mood disturbance score [(F+ T+D+A +C)-V] was scored for analyses because we had no hypotheses about individual mood domains.

Self-rated performance

To assess participants' perceptions of their performance on the morning quiz and GRE tests, they completed ratings of subjective performance, with every point anchored: 'Overall, how would you rate your performance on the test that you just completed?'. Response categories were: 1 = 'very poor'; 2 = 'poor'; 3 = 'good'; 4 = 'very good'; and 5 = 'excellent'.

Hangover

The Acute Hangover Scale (AHS) [64], developed based on empirical hangover data [36,65,66], consists of eight validated symptoms plus 'hangover' rated from 0, 'none' to 7, 'incapacitating' on anchored Likert-type scales. The nine items form a reliable and valid scale, scored using the mean.

Alcohol Administration Manipulation checks

An AlcoSensor-4 (Intoximeters, Inc., St Louis, MO, USA) was used for breath testing. Following beverage adminis-

tration, participants were asked to estimate their blood alcohol concentration on a scale ranging from 0 to 0.15 g%.

Statistical power

With a target enrollment of 200 participants, our study had 99% power of detecting the anticipated medium-sized effect of alcohol on next-day academic test performance ($d = 0.52$), a value derived from our previous studies. For comparison of the effects of alcohol versus placebo in females versus males, the study had 80% power of detecting a difference.

Data analysis approach

All measures were examined for normality and outliers, using the criteria set forth by Hoaglin *et al.* [67]. Outliers were recoded following recommendations by Tabachnick & Fidell [68]. Among the primary outcomes measures, there was one outlier for both the GRE verbal and GRE quantitative scores and five outliers for the quiz score.

Differences in outcomes following consumption of alcohol versus placebo were tested through mixed-effects regression models for repeated-measures data [69]. Our primary interest was in differences by experimental condition (alcohol versus placebo, a within-subjects factor). We controlled for randomly assigned order of beverage administration by including a session variable (indicating a first or second study evening, a within-subject factor) and also controlled for gender (a between-subject factor). Differences in alcohol effects for males and females were tested through the interaction between experimental condition and gender, and all other two-way and three-way interactions were also included in the model. Where significant interactions were found between experimental condition and gender, within-gender alcohol effects were tested through model contrasts.

Comparisonwise *P*-values are reported. When considering multiple testing issues, we grouped study outcomes as measures of: (i) academic performance (one quiz and two GRE scores); (ii) 10 primary neurocognitive performance measures (including the PVT); (iii) nine exploratory neurocognitive performance measures; (iv) mood state measures (a.m. and p.m. assessments); and (v) self-reported performance (one for the quiz and one for the two GRE scores). Analyses are interpreted to indicate an alcohol effect if either the main effect of beverage, or the interaction between experimental condition and gender, are significant. To account formally for multiple comparisons using a Bonferroni adjustment, comparisonwise *P*-values of 0.008 (academics) 0.0025 (primary neurocognitive) 0.0028 (exploratory neurocognitive) and 0.0125 (mood state and self-rated performance) would be required. Because Bonferroni is known to overcorrect, we used an $\alpha = 0.005$ throughout our analyses.

Although formal analyses were based on mixed effects regression models, rather than simple differences by beverage condition, difference scores and their standard deviations are presented for ease of interpretation. Differences in performance are also described as standardized effect sizes, calculated as the difference in mean performance under alcohol and placebo divided by the standard deviation of the difference scores (Cohen's *d*) [70]. Cohen [70] considers effect sizes (*d*) of 0.2, 0.5 and 0.8 as small, moderate and large, respectively.

RESULTS

Participant enrollment

Four hundred and thirteen participants were screened; 364 (88%) were eligible. Of these, 239 (65%) appeared for their scheduled experimental session, and of these 196 (82%) completed the study. Three of the 196 participants who completed the study were excluded from analyses because their maximum breath alcohol measures did not reach the minimum BrAC level (.09 g%). Seventy per cent of participants reported some hangover on the morning following alcohol administration. The mean AHS score was significantly higher under alcohol condition, relative to placebo condition (Table 1).

Objective performance outcomes

The morning after beverage administration, neither the quiz scores on the prior day's lecture nor the two GRE scores differed by beverage condition; effect sizes were close to zero (<0.06). None of the academic performance outcomes showed significant beverage–order or gender–beverage interactions (Table 3).

Of the nine primary NES3 measures, VST-B was significantly different by beverage. PMT-C showed significant gender by beverage interaction ($P = 0.032$); females performed worse (borderline significant) under alcohol condition, relative to placebo, but for males there was no difference. No interactions of beverage with order were significant. The morning after beverage administration, median attention/reaction-time scores, as measured by the PVT, were significantly longer under the alcohol condition, relative to the placebo condition (Table 4). Of the exploratory neurocognitive tests, none was significantly different by beverage condition at our α level.

Dependent measures of subjective effects

Mood

The day after beverage administration, the mean total mood disturbance score was significantly worse under

Table 3 Academic performance outcomes by experimental condition.

	Measure	*n*	*Alcohol*	*Placebo*	*Difference (SD)*	*Effect Size (d)*	*P-value*
GRE Raw Scores	GRE verbal	193	495.39 (87.79)	497.62 (86.43)	−2.23 (61.02)	0.04	NS
	GRE quantitative	193	615.75 (98.92)	612.38 (94.64)	+3.37 (62.57)	0.05	NS
Quiz no. correct		193	24.70 (2.26)	24.59 (2.48)	+0.11 (2.65)	0.04	NS

All *P*-values are based on mixed-effects models controlling for gender and session number. The interaction of gender and dose was tested in each model and found to be non-significant. GRE: Graduate Record Examinations; NS: not significant; SD: standard deviation.

Table 4 Neurobehavioral Evaluation System-3 and PVT outcomes by beverage condition.

NES3 outcomes	*n*	*Alcohol*	*Placebo*	*Difference (SD)*	*Effect size (d)*	*P-value*
Tests requiring speed						
Finger Tapping Test: mean number of taps, preferred hand (FTT-P)	188	59.68 (7.11)	60.12 (7.24)	−0.44 (4.73)	0.09	NS
Sequences Test (ST-A-L)						
Sequence A: latency (ms)[a]	188	14.35 (2.66)	14.48 (3.02)	−0.13 (2.59)	0.05	NS
Digit-Symbol Test (DST-L)						
Latency (ms)[b]	188	80.02 (9.53)	79.53 (9.22)	+0.49 (6.63)	0.07	NS
Pattern Memory Test (PMT-L)						
Average response latency for correct items (seconds)	188	3.17 (0.85)	3.15 (0.90)	+0.01 (0.71)	0.02	NS
Tests requiring sustained attention						
Auditory Digit Span Test (ADST-B)[c]						
Maximum span backward	188	6.25 (1.40)	6.16 (1.42)	+0.09 (1.40)	0.06	NS
Adaptive Paced Auditory Serial Addition Test (APASAT-C)						
Number correct	184	94.92 (3.42)	95.06 (3.19)	−0.14 (2.86)	0.05	NS
Visual Span Test (VST-B)						
Maximum span backward	186	5.41 (0.89)	5.67 (1.16)	−0.26 (1.22)	0.21	0.004
Pattern Memory Test (PMT-C)						
Number correct						
Male	103	16.14 (2.90)	16.06 (2.36)	+0.08 (2.65)	0.03	NS
Female	85	15.26 (2.74)	16.12 (2.12)	−0.86 (2.70)	0.32	0.004
Tests requiring sustained attention and reaction-time						
Continuous performance test (CPT)						
Reaction-time (ms)	187	378.77 (35.48)	375.98 (35.82)	+2.78 (22.47)	0.12	NS
Psychomotor vigilance test (PVT)						
Median reaction-time (ms)	190	223.40 (22.81)	218.57 (20.25)	+4.83 (15.08)	0.32	0.000

All *P*-values are based on mixed-effects models controlling for gender and session number. The interaction of gender and dose was tested in each model. If interaction found to be significant, results were presented by gender. [a]Maximum time permitted to complete sequence A: 60 seconds; sequence B: 120 seconds. [b]Maximum time permitted to complete digit/symbol test: 180 seconds. [c]Valid range of span scores for the forward condition: 3–9; backward condition: 2–8. NS: not significant; SD: standard deviation.

alcohol condition, relative to placebo condition, in both the morning and the afternoon (Table 5).

Self-rated performance

Participants tended to rate their performance on the academic tests as worse under alcohol condition, compared to placebo condition. These differences were significant for self-rated performance on the quiz and GREs (Table 5). Participants' mean estimates of their BrACs following beverage administration were 0.006 g% and 0.098 g% under placebo and alcohol conditions, respectively.

DISCUSSION

College students' test-taking performance was not affected significantly on the morning after intoxication. Significant decrements in some laboratory tests of neu-

Table 5 Subjective measures by beverage condition.

Profile of Mood States (POMS) (higher scores reflect more negative mood state)

Measure	*n*	*Alcohol*	*Placebo*	*Difference (SD)*	*Effect Size (d)*	*P-value*
Morning: total mood disturbance score	193	6.71 (9.41)	1.90 (7.20)	+4.81 (7.95)	0.60	0.000
Afternoon: total mood disturbance score	153	4.30 (10.19)	1.93 (8.39)	+2.37 (8.72)	0.27	0.001
Self-rated performance						
Quiz performance	185	3.43 (0.77)	3.61 (0.79)	−0.18 (0.95)	0.19	0.005
GRE performance	188	2.48 (0.69)	2.65 (0.68)	−0.18 (0.76)	0.23	0.002

All *P*-values are based on mixed-effects models controlling for gender and session number. The interaction of gender and dose was tested in each model and found to be non-significant. GRE: Graduate Record Examinations; SD: standard deviation.

rocognitive function were observed on the morning after alcohol. The NES3 was administered to increase understanding of academic performance effects, should they be found. Under placebo condition, participants' NES3 performance scores were normative and most tests showed no alcohol effects. The pattern of residual alcohol effects we found clustered around visuospatial, motor function and attention/reaction-time deficits. These effects may not be central to performance on multiple choice tests based on recall and recognition, but may affect other types of academic performance (unmeasured by our study), such as essay-writing and problem-solving requiring higher-order cognitive skills, as well as safety-related performance such as ability to process information and respond quickly to unexpected events when driving or operating machinery. Mood states, both in the morning and afternoon, were significantly worse on the day after alcohol. Similarly, participants tended to rate their test-taking performance as significantly worse on the day after alcohol relative to placebo, even though no impairment in academic performance was actually observed.

We do not believe our outcomes were artifacts of participant motivation. The GRE scores were comparable to recent norms, with about 60% of participants scoring in the top 50th percentile of the national distribution. Similarly, the mean quiz scores were about 83%, high enough to indicate participant motivation, but low enough to suggest that the quizzes were not too easy (i.e. no ceiling effect). We also do not believe that participant blinding, which can be problematic at high alcohol doses, affected results because the bias would be away from the null hypothesis and we did not find differences on the primary outcome variables (academic test-taking performance). Although our procedures called for abstinence from recreational drugs 24 hours prior to experimental sessions, we used only self-report to check drug-use compliance. Moreover, we did not screen for, or document, drug-use history. Thus, participants' undisclosed drug use prior to experimental sessions could have. If so, there was no consistent effect, as some outcomes were affected significantly on the day after alcohol and others were not.

Although the morning and afternoon mood scores were significantly worse following the alcohol condition, these results may have been driven in part by fatigue resulting from alcohol's sleep-disturbing effects [36,71–73].

While our findings are discordant with results of survey studies that find associations between alcohol use and academic problems, these studies are potentially confounded in that a third factor (e.g. personality) may cause both excessive drinking and academic difficulties and causal order is unknown (i.e. academic difficulties could lead to excessive drinking). Our findings are consistent, however, with a study on the effects of intoxication on next-day occupational performance [33]. In that study, merchant marine cadets' performance on a diesel engine simulator was not affected significantly, relative to placebo, on the morning after intoxication (mean BrAC.115 g%), but self-rated performance was significantly worse. Similarly, another laboratory study found measures of combined attention and reaction-time to be the only neurocognitive measures affected on the morning after 0.11 g% BrAC [74].

We do not conclude, however, that excessive drinking is not a risk factor for academic problems. It is possible that a higher alcohol dose would have affected next-day academic test scores. Moreover, test-taking is only one factor in academic success. Study habits, motivation and class attendance also contribute to academic performance; each of these could be affected by intoxication. When drinking leads to staying up too late, sleeping in or getting too little sleep, it can disrupt next-morning attendance or focus. Moreover, we did not measure whether learning skills were impaired on the day after intoxication. The neurocognitive measures that were affected negatively on the day after alcohol could be related to the ability to process new information effectively. By neces-

sity, all participants were ≥21 years of age and thus were college juniors, seniors or recent graduates. It is possible that over the course of their education students develop skills that allow them to perform well on multiple-choice tests despite neurocognitive impairment resulting from intoxication the previous night. Accordingly, had our participants been freshmen or sophomores, they might have performed worse under alcohol, relative to placebo, condition. We excluded volunteers who had not engaged in recent binge drinking or who were at risk for alcohol dependence. It is possible that these excluded drinkers might be more susceptible to alcohol-related problems with test-taking. Nonetheless, in surveys almost half of college students report binge drinking and presumably most of these have not developed alcohol dependence. Thus, we believe that our findings are relevant to a substantial proportion of college students.

Clinical trials registration

ClinicalTrials.Gov Identifier: NCT00183170

Declarations of interest

None.

Acknowledgements

We wish to acknowledge the invaluable contribution of Daniel J. Gottlieb, MD, who provided clinical support and expertise to this study. This research was supported by: (1) the Youth Alcohol Prevention Center, Boston University School of Public Health, with funding from NIAAA grant # P60 AA013759-01; and (2) by the National Center for Research Resources (NCRR), grant # M01 RR00533, a component of the National Institutes of Health (NIH). Its contents are solely the responsibility of the authors and do not necessarily represent the official view of the NCRR or NIH.

References

1. National Advisory Council on Alcohol Abuse and Alcoholism. *Recommended Council Guidelines on Ethyl Alcohol Administration in Human Experimentation*. Revised May 2005. Available at: http://www.niaaa.nih.gov/Resources/Research Resources (accessed 12 March 2009).
2. Chen C. M., Dufour M. C., Yi H.-Y. Alcohol consumption among young adults ages 18–24 in the United States: results from the 2001–2002 NESARC Survey. *Alcohol Res Health* 2004/2005; **28**: 269–80.
3. White H. R., Jackson K. Social and psychological influences on emerging adult drinking. *Alcohol Res Health* 2004/2005; **28**: 182–90.
4. Wechsler H., Nelson T. F. What have we learned from the Harvard School of Public Health College Alcohol Study: focusing attention on college student alcohol consumption and environmental conditions that promote it? *J Stud Alcohol Drugs* 2008; **69**: 481–90.
5. Slutske W. S., Piasecki T. M., Hunt-Carter E. E. Development and initial validation of the Hangover Symptom Scale: prevalence and correlates of hangover symptoms in college students. *Alcohol Clin Exp Res* 2003; **27**: 1442–50.
6. Hingson R. W., Heeren T., Winter M., Wechsler H. Magnitude of alcohol-related mortality and morbidity among US college students ages 18–24: changes from 1998–2001. *Annu Rev Public Health* 2005; **26**: 259–79.
7. Perkins H. W. Surveying the damage: a review of research on consequences of alcohol misuse in college populations. *J Stud Alcohol* 2002; Suppl **14**: 91–100.
8. Atkinson R. C., Shiffrin R. M. Human memory: a proposed system and its control processes. In: Spence K. W., Spence J. T., editors. *The Psychology of Learning and Motivation*, London: Academic Press; 1968, p. 90–191.
9. Wechsler H., Dowdall G. W., Maenner G., Gledhill-Hoyt J., Lee H. Changes in binge drinking and related problems among American college students between 1993 and 1997: results of the Harvard School of Public Health College Alcohol Study. *J Am Coll Health* 1998; **47**: 57–68.
10. Perkins H. W. Gender patterns in consequences of collegiate alcohol abuse: a 10-year study of trends in an undergraduate population. *J Stud Alcohol* 1992; **53**: 458–62.
11. Werch C. E., Gornam D. R., Marty P. J. The relationship between alcohol consumption and alcohol problems in young adults. *J School Health* 1987; **57**: 232–6.
12. Wolaver A. M. Effects of heavy drinking on study effort, grade point average, and major choice. *Contemp Econ Policy* 2002; **20**: 415–28.
13. Engs R. C., Hanson D. J., Diebold B. A. The drinking patterns and problems of a national sample of college students. *J Alcohol Drug Educ* 1996; **41**: 13–33.
14. Singleton R. A. Collegiate alcohol consumption and academic performance. *J Stud Alcohol Drugs* 2007; **68**: 548–55.
15. Singleton R. A., Wolfson A. R. Alcohol consumption, sleep, and academic performance among college students. *J Stud Alcohol Drugs* 2009; **70**: 355–63.
16. Wood P. K., Sher K. J., Erickson D. J., DeBord K. A. Predicting academic problems in college from freshman alcohol involvement. *J Stud Alcohol Drugs* 1997; **58**: 200–10.
17. Paschall M. J., Freisthler B. Does heavy drinking affect academic performance in college? Findings from a prospective study of high achievers. *J Stud Alcohol Drugs* 2003; **64**: 515–9.
18. Wolkenberg R., Gold C., Tichauer E. Delayed effects of acute alcoholic intoxication on performance with reference to work safety. *J Safety Res* 1975; **7**: 104–19.
19. Seppälä T., Leino T., Linnoila M., Huttunen M., Ylikahri R. Effects of hangover on psychomotor skills related to driving: modification by fructose and glucose. *Acta Pharmacol Toxicol* 1976; **38**: 209–18.
20. Laurell H., Törnros J. Investigation of alcoholic hangover effects on driving performance. *Blutalkohol* 1983; **20**: 489–99.
21. Törnros J., Laurell H. Acute and hang-over effects of alcohol on simulated driving performance. *Blutalkohol* 1991; **28**: 24–30.
22. Yesavage J., Leirer V. Hangover effects on aircraft pilots 14 hours after alcohol ingestion: a preliminary report. *Am J Psychiatry* 1986; **143**: 1546–50.
23. Morrow D., Leirer V., Yesavage J. The influence of alcohol and aging on radio communication during flight. *Aviat Space Environ Med* 1990; **61**: 12–20.
24. Morrow D., Leirer V., Yesavage J., Tinklenberg J. Alcohol,

age, and piloting: judgment, mood, and actual performance. *Int J Addict* 1991; **26**: 669–83.
25. Morrow D., Yesavage J., Leirer V., Dohlert N., Taylor J., Tinkleberg J. The time-course of alcohol impairment of general aviation pilot performance in a Frasca 141 simulator. *Aviat Space Environ Med* 1993; **64**: 697–705.
26. Taylor J., Dohlert N., Morrow D., Friedman L., Yesavage J. Acute and 8-hour effects of alcohol (0.08% BAC) in younger and older pilots' simulator performance. *Aviat Space Environ Med* 1994; **65**: 718–25.
27. Taylor J. L., Dolhert N., Friedman L., Mumenthaler M., Yesavage J. A. Alcohol elimination and simulator performance of male and female aviators: a preliminary report. *Aviat Space Environ Med* 1996; **67**: 407–13.
28. Yesavage J., Dolhert N., Taylor J. Flight simulator performance of younger and older aircraft pilots: effects of age and alcohol. *J Am Geriatr Soc* 1994; **42**: 577–82.
29. Petros T., Bridewell J., Jensen W., Ferraro F. R., Bates J. A., Moulton P. *et al.* Postintoxication effects of alcohol on flight performance after moderate and high blood alcohol levels. *Int J Aviat Psychol* 2003; **13**: 287–300.
30. Dowd P. J., Wolfe J. W., Cramer R. L. After effects of alcohol on the perception and control of pitch attitude during centripetal acceleration. *Aerosp Med* 1973; **44**: 928–30.
31. Collins W. Performance effects of alcohol intoxication and hangover at ground level and at simulated altitude. *Aviat Space Environ Med* 1980; **51**: 327–35.
32. Collins W., Chiles W. Laboratory performance during acute alcohol intoxication and hangover. *Hum Factors* 1980; **22**: 445–62.
33. Rohsenow D. J., Howland J., Minsky S., Arnedt J. T. Effects of heavy drinking by maritime academy cadets on hangover, perceived sleep, and next-day ship power plant operation. *J Stud Alcohol* 2006; **67**: 405–15.
34. Streufert S., Pogash R., Braig D., Gingrich D., Kantner A., Landis R. *et al.* Alcohol hangover and managerial effectiveness. *Alcohol Clin Exp Res* 1995; **19**: 1141–6.
35. Myrsten A. L., Rydberg U., Idelström C. M., Lamble R. Alcohol intoxication and hangover: modification of hangover by chlormethiazole. *Psychopharmacology* 1980; **69**: 117–25.
36. Roehrs T., Yoon J., Roth T. Nocturnal and next-day effects of ethanol and basal level of sleepiness. *Hum Psychopharmacol* 1991; **6**: 307–11.
37. Roehrs T., Roth T. Sleep, sleepiness, and alcohol use. *Alcohol Res Health* 2001; **25**: 101–9.
38. Roehrs T., Roth T. Sleep, sleepiness, sleep disorders and alcohol use and abuse. *Sleep Med Rev* 2001; **5**: 287–97.
39. Verster J. C., van Duin D., Volkerts E., Schreude A. H., Verbaten M. N. Alcohol hangover effects on memory functioning and vigilance performance after an evening of binge drinking. *Neuropsychopharmacology* 2003; **28**: 740–6.
40. Kim D. J., Yoon S. J., Lee H. P., Choi B. M., Go H. J. The effects of alcohol hangover on cognitive functions in healthy subjects. *Int J Neurosci* 2003; **113**: 581–94.
41. McKinney A., Coyle K. Alcohol hangover effects on measures of affect the morning after a normal night's drinking. *Alcohol Alcohol* 2006; **41**: 54–60.
42. Alford C., Wadling S. Comparative effects of caffeine and alcohol bedtime drinks on sleep, performance and mood in young adults. *J Psychopharmacol* 2004; **18**: A41.
43. Finnigan F., Schulze D., Smallwood J., Helander A. The effects of self-administered alcohol-induced 'hangover' in a naturalistic setting on psychomotor and cognitive performance and subjective scale. *Addiction* 2005; **100**: 1680–9.
44. Kruisselbrink L. D., Martin K. L., Megeney M., Fowles J. R., Murphy R. J. Physical and psychomotor functioning of females the morning after consuming low to moderate quantities of beer. *J Stud Alcohol* 2006; **67**: 416–20.
45. Takala M., Siro E., Toivainen Y. Intellectual functions and dexterity during hangover-experiments after intoxication with brandy and with beer. *Q J Stud Alcohol* 1958; **19**: 1–29.
46. McCaul M. E., Turkkan J. S., Svikis D. S., Bigalow G. E. Alcohol and serobarbital effects as a function of familial alcoholism: extended intoxication and increased withdrawal effects. *Alcohol Clin Exp Res* 1991; **15**: 94–101.
47. Lemon J., Chesher G., Fox A., Greeley J., Nabke C. Investigation of the 'hangover' effects of an acute dose of alcohol on psychomotor performance. *Alcohol Clin Exp Res* 1993; **17**: 665–8.
48. Chait L. D., Perry J. L. Acute and residual effects of alcohol and marijuana, alone and in combination, on mood and performance. *Psychopharmacology* 1994; **115**: 340–9.
49. Finnegan F., Hammersley R., Cooper T. An examination of next-day hangover effects after a 100 mg/100 ml dose of alcohol in heavy social drinkers. *Addiction* 1998; **93**: 1829–38.
50. Selzer M. D., Vinokur A., Van Rooijen L. A self-administered short Michigan alcoholism screening tests (SMAST). *J Stud Alcohol* 1975; **36**: 117–26.
51. Brick J., Nathan P. E., Westrick E., Frankenstein W., Shapiro A. The effects of menstrual cycle on blood alcohol levels and behavior. *J Stud Alcohol* 1986; **47**: 472–7.
52. Niaura R. S., Nathan P. E., Frankenstein W., Shapiro A. P., Brick J. Gender differences in acute psychomotor, cognitive, and pharmacokinetic response to alcohol. *Addict Behav* 1987; **12**: 345–56.
53. Terner J. M., deWit H. Menstrual cycle phase and responses to drugs of abuse in humans. *Drug Alcohol Depend* 2006; **84**: 1–13.
54. Friel P. N., Logan B. K., O'Malley D., Baer J. S. Development of dosing guidelines for reaching selected target breath alcohol concentrations. *J Stud Alcohol* 1999; **60**: 555–65.
55. Keane T. M., Lisman S. A., Kreutzer J. Alcoholic beverages and their placebos: an empirical evaluation of expectancies. *Addict Behav* 1980; **5**: 313–28.
56. Tassi P., Muzet A. Sleep inertia. *Sleep Med Rev* 2000; **4**: 341–53.
57. Mann R. E., Sobell L. C., Sobell M. D., Pavan D. Reliability of a family tree questionnaire for assessing family history of alcohol problems. *Drug Alcohol Depend* 1985; **15**: 61–7.
58. Schneider M. J. *Introduction to Public Health*. Boston, MA: Jones and Barrlett Publishers; 2006.
59. Educational Testing Service. *Preparing for the GRE General Test* [Online]. 2001. 24 May. Available at: http://www.gre.org/codelst.html (accessed 15 September 2001).
60. White R. F., James K. E., Vasterling J. J., Letz R., Marans K., Delaney R. *et al.* Neuropsychological screening for cognitive impairment using computer-assisted tasks. *Assessment* 2003; **10**: 86–101.
61. Lezak M. D. *Neuropsychological Assessment*, 2nd edn. New York: Oxford Press; 1983.
62. Dinges D. F., Powell J. E. Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. *Behav Res Methods Instrum Comput* 1985; **17**: 652–5.

63. McNair D. M., Lorr M., Droppleman L. F. *Profile of Mood States: POMS Manual*. San Diego, CA: EdITS; 1992.
64. Rohsenow D. J., Howland J., Minsky S. J., Greece J., Almeida A., Roehrs T. The acute hangover scale: a new measure of immediate hangover symptoms. *Addict Behav* 2007; **32**: 1314–20.
65. Chapman L. Experimental induction of hangover. *Q J Stud Alcohol* 1970; **5**(Suppl 5): 67–86.
66. Ylikahri R., Huttumen M., Eriksson C. J. P., Hikkila E. A. Metabolic studies on the pathogenesis of hangover. *Eur J Clin Invest* 1974; **4**: 93–100.
67. Hoaglin D., Mosteller F., Tukey J. *Understanding Robust and Exploratory Data Analysis*. New York: John Wiley & Sons; 1983.
68. Tabachnick B., Fidell L. *Using Multivariate Statistics*. New York: Harper & Row; 1983.
69. Diggle P. J., Liang K. Y., Zeger S. L. *Analysis of Longitudinal Data*. Oxford: Oxford University Press; 1994.
70. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd edn. Mahwah, NJ: Lawrence Erlbaum; 1988.
71. Stone B. M. Sleep and low doses of alcohol. *Electroencephalogr Clin Neurophysiol* 1980; **48**: 706–9.
72. Kobayashi T., Misaki K., Nakagawa H., Okuda K., Ota T., Kanda I. *et al.* Alcohol effect on sleep electroencephalography by fast Fourier transformation. *Psychiatry Clin Neurosci* 1998; **52**: 154–5.
73. Williams D. L., MacLean A. W., Cairns J. Dose–response effects on ethanol on the sleep of young women. *J Stud Alcohol* 1983; **44**: 515–23.
74. Rohsenow D. J., Howland J., Arnedt J. T., Almeida A. B., Greece J. A., Minsky S. *et al.* Intoxication with bourbon versus vodka: effects on hangover, sleep and next-day neurocognitive performance in young adults. *Alcohol Clin Exp Res* 2009. Epub ahead of print 17 December 2009, DOI: 10.1111/j1530-0277.2009.01116.x

EDITORIAL

Therapeutic Potential of Oral Factor Xa Inhibitors

Elaine M. Hylek, M.D., M.P.H.

Venous thromboembolism is the third leading cause of cardiovascular death, after myocardial infarction and stroke.[1] Total hip or knee arthroplasty is the procedure with the highest risk of venous thromboembolism.[2] In this issue of the *Journal*, two studies affirm and extend the efficacy and safety of the novel oral factor Xa inhibitors, rivaroxaban and apixaban, in the management of venous thromboembolic disease.[3,4] In the Acute DVT Study,[3] rivaroxaban (at a dose of 15 mg twice daily for 3 weeks, followed by 20 mg once daily) was compared with enoxaparin followed by warfarin or acenocoumarol, for 3, 6, or 12 months, in patients with acute, symptomatic deep-vein thrombosis. Rivaroxaban had noninferior efficacy with respect to recurrent venous thromboembolism, with similar rates of hemorrhage. The Continued Treatment Study[3] confirmed the persistent risk of recurrent venous thromboembolism after initial treatment, as shown by Ridker and colleagues,[5] and lends further support to extending the duration of anticoagulant therapy, particularly given the low rates of major bleeding with rivaroxaban (0.7%). Lassen and colleagues studied thromboprophylactic regimens in patients undergoing total hip replacement.[4] Participants were randomly assigned to apixaban, at a dose of 2.5 mg orally twice daily, or enoxaparin, at a dose of 40 mg subcutaneously every 24 hours, with the treatments initiated perioperatively and continued for 35 days after surgery. Apixaban was associated with lower rates of venous thromboembolism without an increase in bleeding complications.

The oral factor Xa inhibitors represent a major advance in the prevention and treatment of thromboembolic disease. Factor Xa is strategically positioned at the juncture of the intrinsic and extrinsic coagulation pathways proximal to thrombin. The potential impact of these oral, highly specific, fixed-dose drugs that do not require routine monitoring will no doubt be substantial. Currently, millions of people worldwide are relegated to receiving no therapy or therapy that has been proven to be ineffective, because they lack access to the monitoring expertise needed to safely and effectively administer warfarin. It is conceivable that the oral factor Xa inhibitors, as compared with warfarin, will prove to be safer in clinical practice because they are administered in fixed doses, do not interfere with diet, and have fewer interactions with other drugs. Given the nine different tablet strengths of warfarin, transitions in care settings and fluctuations in health status invariably create opportunities for unintended harm. A growing appreciation of the hazards of warfarin therapy prompted the Food and Drug Administration to issue a black-box warning for warfarin in October 2006.[6] The factor Xa inhibitors that are most advanced in clinical development are rivaroxaban, apixaban, and edoxaban. (Other factor Xa inhibitors in development include betrixaban, YM150, and TAK-442.) Unlike the case with warfarin, drug elimination in the case of the factor Xa inhibitors involves multiple pathways. The degree of renal clearance is 66% in the case of rivaroxaban, 25% in the case of apixaban, and 35% in the case of edoxaban. As compared with warfarin's half-life of 20 to 60 hours, the respective half-lives of these agents are 7 to 11 hours, 12 hours, and 9 to 11 hours. As shown in the study by the EINSTEIN investigators, the rapid onset of action obviates the need for heparin in the acute management of venous thrombosis. The rapid onset of action also has impli-

cations for the appropriate timing of the initiation of the drug after the procedure, given the need for wound hemostasis. The shorter half-life of these agents may improve their overall safety profile but, conversely, will also result in the drugs' providing less protection if doses are missed. All these drugs are metabolized to different degrees by cytochrome P-450 3A4 (CYP3A4) and are substrates for P-glycoprotein. Therefore, the concomitant use of drugs that inhibit both pathways, such as azole antifungal agents or protease inhibitors, is contraindicated.[7]

Each of the factor Xa inhibitors is being evaluated in at least one large-scale phase 3 trial of stroke prevention in patients with atrial fibrillation. Positive results from the completed 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; ClinicalTrials.gov number, NCT00403767) were recently presented at the annual scientific sessions of the American Heart Association.[8] The randomized trial of Apixaban versus Acetylsalicylic Acid to Prevent Stroke in Atrial Fibrillation Patients Who Have Failed or Are Unsuitable for Vitamin K Antagonist Treatment (AVERROES, NCT00496769) was stopped early because the efficacy of apixaban had been shown.[9] The results of ongoing trials involving patients with atrial fibrillation, in which warfarin is the active comparator, are expected in 2011: the Apixaban for the Prevention of Stroke in Subjects With Atrial Fibrillation study (ARISTOTLE, NCT00412984), in which apixaban is being tested with a dose of 5 mg twice daily, and the Global Study to Assess the Safety and Effectiveness of DU-176b versus Standard Practice of Dosing With Warfarin in Patients With Atrial Fibrillation (EngageAFTIMI48, NCT00781391), in which two doses of edoxaban, 30 mg and 60 mg, each administered once daily, are being compared with warfarin.

Translating the efficacy and safety that have been shown in clinical trials to real-world practice is often a challenge because, as compared with patients in real-world practices, participants in trials are usually younger, have less medically complex illnesses, are more likely to be adherent, and have been specifically selected on the basis of having a lower risk of bleeding. Concomitant antiplatelet therapy is either discouraged or considered to be an exclusion criterion. The mean age of participants undergoing hip arthroplasty in the study by Lassen et al. was 60 years, and approximately 89% of the participants had normal renal function. Similarly, the mean age of participants with acute symptomatic deep-vein thrombosis in the study by the EINSTEIN investigators was 56 years, and 92% had a creatinine clearance of 50 ml per minute or more. Because both the risk of thrombosis and the risk of hemorrhage increase substantially with age and with burden of chronic disease, the effectiveness of the novel agents in real-world practice will need to be closely monitored, particularly among older adults with renal impairment. The critical role of baseline risk and the additive hazards of combination antiplatelet therapy and bleeding were highlighted by the recent early termination, because of increased bleeding with apixaban, of the Apixaban for Prevention of Acute Ischemic Events 2 trial (APPRAISE-2, NCT00831441), in which patients with a recent acute coronary syndrome who were receiving single or dual antiplatelet therapy were randomly assigned to apixaban or placebo.[10]

Alternatives to warfarin have been long awaited. The oral factor Xa inhibitors show great promise. The reversibility of the drugs' effects and the ability to measure the anticoagulant effect in specific situations will continue to be highly desirable features and will help to allay physicians' concerns. If these novel, breakthrough, oral anticoagulant drugs prove to be effective across the broad spectrum of patients in routine care and are conscientiously priced, the worldwide impact will be huge.

Disclosure forms provided by the author are available with the full text of this article at NEJM.org.

From the Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, Boston.

1. Anderson FA Jr, Wheeler HB. Physician practices in the management of venous thromboembolism: a community-wide survey. J Vasc Surg 1992;16:707-14.
2. White RH. The epidemiology of venous thromboembolism. Circulation 2003;107:Suppl 1:I4-I8.
3. The EINSTEIN Investigators. Oral rivaroxaban for symptomatic venous thromboembolism. N Engl J Med 2010;363:2499-510.
4. Lassen MR, Gallus A, Raskob GE, Pineo G, Chen D, Ramirez LM. Apixaban versus enoxaparin for thromboprophylaxis after hip replacement. N Engl J Med 2010;363:2487-98.
5. Ridker PM, Goldhaber SZ, Danielson E, et al. Long-term, low-intensity warfarin therapy for the prevention of recurrent venous thromboembolism. N Engl J Med 2003;348:1425-34.
6. MedWatch safety alerts: Coumadin tablets (warfarin sodium tablets, USP) crystalline. Silver Spring, MD: Food and Drug Ad-

ministration, 2006. (http://www.fda.gov/medwatch/safety/2006/coumadin_medguide.pdf.)
7. 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.
8. Mahaffey KW, Fox KAA. 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). (http://sciencenews.myamericanheart.org/pdfs/ROCKET_AF_pslides.pdf.)
9. AVERROES study of investigational agent apixaban closes early due to clear evidence of efficacy. BusinessWire. June 10, 2010. (http://www.businesswire.com/news/home/20100610006606/en/AVERROES-Study-Investigational-Agent-Apixaban-Closes-Early.)
10. APPRAISE-2 study with investigational compound apixaban in acute coronary syndrome discontinued. BusinessWire. November 18, 2010. (www.businesswire.com/news/bms/20101118007161/en/APPRAISE-2-Study-Investigational-Compound-Apixaban-Acute-Coronary.)

Jackson et al. BMC Medical Education 2010, 10:22
http://www.biomedcentral.com/1472-6920/10/22

BMC Medical Education

DEBATE Open Access

Internal medicine residency training for unhealthy alcohol and other drug use: recommendations for curriculum design

Angela H Jackson[1,2,3*], Daniel P Alford[1,2,3], Catherine E Dubé[3], Richard Saitz[1,2,3,4,5]

Abstract

Background: Unhealthy substance use is the spectrum from use that risks harm, to use associated with problems, to the diagnosable conditions of substance abuse and dependence, often referred to as substance abuse disorders. Despite the prevalence and impact of unhealthy substance use, medical education in this area remains lacking, not providing physicians with the necessary expertise to effectively address one of the most common and costly health conditions. Medical educators have begun to address the need for physician training in unhealthy substance use, and formal curricula have been developed and evaluated, though broad integration into busy residency curricula remains a challenge.

Discussion: We review the development of unhealthy substance use related competencies, and describe a curriculum in unhealthy substance use that integrates these competencies into internal medicine resident physician training. We outline strategies to facilitate adoption of such curricula by the residency programs. This paper provides an outline for the actual implementation of the curriculum within the structure of a training program, with examples using common teaching venues. We describe and link the content to the core competencies mandated by the Accreditation Council for Graduate Medical Education, the formal accrediting body for residency training programs in the United States. Specific topics are recommended, with suggestions on how to integrate such teaching into existing internal medicine residency training program curricula.

Summary: Given the burden of disease and effective interventions available that can be delivered by internal medicine physicians, teaching about unhealthy substance use must be incorporated into internal medicine residency training, and can be done within existing teaching venues.

Background

Unhealthy substance use (SU) is the spectrum from use that risks harm, to use associated with consequences or problems, to the diagnosable conditions substance abuse and dependence often referred to as substance use disorders [1]. Unhealthy SU is a major public health problem in the United States. Many physician interventions (e.g., brief counseling, pharmacotherapy) have proven efficacy. Internal medicine physicians are among the most commonly visited physicians in the US [2]. Yet internal medicine physician training in substance use-related preventive services, diagnosis, treatment, and chronic disease management has been inadequate. This inadequacy leaves patients and the health system without sufficient expertise to address one of the most common and costly health conditions.

Among people 12 and older, there were 20.4 million current users of illicit drugs, 125 million users of alcohol and 72.9 million users of tobacco products, according to the 2006 National Survey on Drug Use andHealth [3]. Of those, 22.6 million alcohol and illicit drug users (9.2% of the population 12 and older) met criteria for substance abuse or dependence. Drug abuse was responsible in 2002 for approximately 26,000 deaths and cost society $180.8 billion [4]. Alcohol use cost society similarly and was responsible for 85,000 deaths [5,6]. Comparatively, coronary heart disease, the leading cause of death in the United States for the past 80 years and a major cause of disability, cost an estimated $151.6 billion in 2007 [7,8].

* Correspondence: angela.jackson@bmc.org
[1]Residency Training Program in Internal Medicine, Boston University School of Medicine, Boston, MA, USA

Education about unhealthy SU is inadequate in medical training [9]. This deficiency persists despite the contribution of SU to disability and premature death, [10] and its prevalence and societal costs [6,11,12]. Screening and management of SU merits a position in medical curricula that reflects its importance and characteristics as a mainstream medical condition [13-17]. Although screening and brief intervention for unhealthy alcohol use is among the most effective and cost-effective preventive services delivered by physicians, its actual delivery is the lowest among comparably ranked services (most often not delivered to those eligible) [18-20].

Many physicians fail to address SU conditions due to discomfort with SU-related patient discussions, [21] deficient knowledge and clinical skills, [22,23] and negative attitudes, [24,25] all resulting in barriers to providing optimal medical care for their patients and reducing the consequences that affect their families and society [26]. The diagnosis of substance abuse or dependence is often missed by physicians and even when the diagnosis is made, many physicians do not know how to respond appropriately using brief intervention or developing an organized plan for referral or treatment and follow-up. While there are many reasons physicians are not performing screening and brief intervention, such as stigma or lack of skill, there may be few local referral resources for patients with SU, once identified. At minimum, the basic clinical skills of screening, assessment, diagnosis, negotiating treatment and ongoing monitoring in SU must be addressed in physician training. These are skills that physicians already routinely employ in the prevention and management of other chronic conditions [27]. SU conditions can be serious and chronic, and risk factors and earlier stage unhealthy use can be recognized, highlighting the need for physicians to embrace their role in preventing, identifying and managing patients with unhealthy SU [28].

Discussion

Physician education

Medical educators have started addressing the need for physician training in unhealthy SU screening, assessment, and management [29-34]. Formal curricula on these subjects have been developed [35,36] and evaluated [37,38] and recommendations for the medical care of addicted patients have been published [13,39,40] Nonetheless, dissemination of up-to-date addiction research and clinical recommendations into physician practice and residency curricula remains a significant challenge [41,42].

Unhealthy substance use education aimed at improving residents' attitudes and clinical practice behaviors has been shown to be effective [43,44]. When residents feel responsible for caring for patients with SU conditions (i.e., "role responsibility"), they develop greater confidence in their ability to screen and refer patients [45]. Wider implementation of known effective clinical practices for addressing SU conditions requires creative strategies to develop a workforce that sees the management of SU conditions as part of its overall mission, is knowledgeable about state-of-the-art approaches to patient management, and is motivated to implement such practices in a range of clinical settings [9,38,46,47]. As noted in the Institute of Medicine Report *Improving the Quality of Health Care for Mental and Substance-Use Conditions*, [26] medical educators have not adequately addressed past recommendations to update training of medical professionals, leaving trainees ill equipped in their ability to care for patients with SU conditions.

The need to implement SU curricula is also supported by the existence of several national initiatives regarding SU care in medical settings–the Joint Commission, which is considering SU-related performance measures for hospitals, a performance measure for alcohol screening included by the Center for Medicare and Medicaid Services in 2009, and national Screening Brief Intervention Referral and Treatment programs supported by federal grants to a number of states in the US [48-50].

Strategies for educational change

There are several strategies that may be employed to foster the adoption of core addiction medicine competencies into mainstream of graduate medical education curricula, each with strengths and limitations. Examples include:

1) **Modifying residency training to support the development of core skills and behaviors** by the program graduates, though residency programs may be reluctant to add new training initiatives to their busy schedules;

2) **Disseminating models for understanding SU conditions that are already familiar to physicians**, for example, highlighting that SU conditions are often chronic diseases with periods of remission and relapse for many patients; [51]

3) **Addressing attitudes towards unhealthy SU and patients with these conditions**, recognizing that attitudinal issues play a large role in physicians' willingness to address SU conditions in their patients. For example, clinical guidelines and protocols may be more readily accepted if championed by opinion leaders and role models who are trusted sources of clinical information (often requiring them to be from the same specialty and profession), effective presenters of new information about changes in clinical practice and viewed as mentors by colleagues and younger trainees;

4) **Recommendations and requirements of accreditation bodies** to serve as catalysts and ultimately for enforcement of change within training programs.

Accreditation and certification to improve resident physician unhealthy substance use education

Academic institutions provide learners with opportunities to develop knowledge and skills that are prerequisites for safe, effective, and competent practice. Accrediting organizations assess educational *programs* to determine whether their content is designed to produce fully competent graduates. Accreditation is granted to those programs meeting their standards. The Accreditation Council for Graduate Medical Education (ACGME) is a private, non-profit council that evaluates and accredits medical residency programs in the United States. The ACGME was established in 1981 based on a consensus in the academic medical community for an independent accrediting organization. Its forerunner was the Liaison Committee for Graduate Medical Education (LCGME) and had been established in 1972.

The mission of the ACGME is to improve health care by assessing and advancing the quality of resident physicians' education through accreditation. For each medical specialty, the ACGME has a Residency Review Committee (RRC) comprised of 6 to 15 volunteer physicians. Members of the residency review committees are appointed by the American Medical Association (AMA) Council on Medical Education and the appropriate medical specialty boards and organizations.

In the evaluation of graduate medical education, the ACGME has shifted from a *descriptive model* focused on structure and measurement of a program's "potential" to train competent physicians, to a model that measures actual training *outcomes*. In 1997, the ACGME initiated the Outcome Project and began to develop core competencies. The goal of the Outcomes Project is to enhance residency education through resident outcome assessment [52]. This project is a long-term initiative which emphasizes the attainment of a core set of competencies by the residents, as an indicator of a residency program's educational effectiveness and quality rather than simple compliance with regulations. In 1999, the AGGME endorsed six general competencies around which all residency curricula should be organized: 1) Medical Knowledge; 2) Patient Care; 3) Interpersonal and Communication Skills; 4) Professionalism; 5) Practice-based learning and improvement; 6) Systems-based practice. The ACGME has progressively moved to the present mandate for training programs to demonstrate data-driven changes and improvements in curricula based on resident performance data in each of the competencies, promoting continuous improvement in resident education and ultimately, in the healthcare workforce. Any efforts to improve resident physician unhealthy substance use education via accreditation will likely be most successful if they take into account and relate clearly to the ACGME core competencies.

Health professional organizations frequently rely on independent certifying bodies that grant certification recognizing that individuals have successfully demonstrated knowledge or competency in a particular specialty. The American Board of Internal Medicine (ABIM) is a non-profit, independent evaluation organization that, through the administration of a certifying examination, has for more than 70 years maintained the highest standard in internal medicine. ABIM certification has meant that internists have demonstrated - to their peers and to the public - that they have the clinical judgment, skills and attitudes essential for the delivery of excellent patient care. However, only 2% of the American Board of Internal Medicine certifying exam typically addresses substance use, which translates into 2-5 questions in the entire exam (compared to 14% for cardiovascular disease, 6% for nephrology, 2% for ophthalmology, and 10% for geriatrics). Regulatory bodies may provide some leverage in instituting more global implementation of resident training in substance abuse conditions by increasing the emphasis of substance use and related conditions on their examinations, such as on the ABIM certifying examination.

Unhealthy substance use-related competencies

The Health Resources and Services Administration (HRSA) and the Substance Abuse and Mental Health Services Administration (SAMHSA) Center for Substance Abuse Treatment (CSAT) supported an effort by the Association for Medical Education and Research in Substance Abuse (AMERSA) to implement an interdisciplinary project to improve health professional education in substance abuse [53]. The project was known as Project MAINSTREAM (the Multi-Agency INitiative on Substance abuse Training and Education for America).

A major aim of this project was to produce a national strategic plan to improve care for substance use problems, including state-of-the-art reviews and recommendations for health professional development by leading authorities. To develop the strategic plan, nationally recognized experts were invited to join a Strategic Planning Advisory Committee (SPAC) representing dentists, dietitians, nurse midwives, nurses, nurse practitioners, occupational therapists, pharmacists, physical therapists, physicians, physician assistants, psychologists, public health professionals, rehabilitation counselors, social workers, speech pathologists, and audiologists.

Using a modified consensus-development approach, they defined a set of core competencies for all health professionals, irrespective of discipline. In addition,

members of the SPAC, in conjunction with other national leaders in substance abuse, developed discipline-specific papers that summarize the state of the art regarding education of health professionals about SU conditions and provide recommendations and action steps for achieving desired goals within each discipline. All of the papers were subjected to peer review and were modified before being accepted for inclusion in the Strategic Plan. Following further review of the papers, an exhaustive stratification process was used to derive key recommendations that cut across the professional disciplines represented by the authors. The recommendations represent the collective input from SPAC members and outside experts from all of the disciplines and hundreds of other individuals who assisted in the review of materials in the Strategic Plan.

Since publication of the Strategic Plan in 2002, recommended physician competencies were adopted by the White House Office of National Drug Control Policy and by medical education leaders (including representatives from the ACGME, AMA, and the Society of General Internal Medicine) in a series of Leadership Conferences on Medical Education in Substance Abuse that took place in 2004, 2006, and twice in 2008, [28]

Unhealthy Substance Use Curriculum for Internal Medicine Residency Programs

Introduction to Unhealthy Substance Use Curriculum Many previous publications have outlined curricula for physicians at various stages of training and from various specialties, for medical schools and residencies. In this paper, we outline a curriculum in unhealthy substance use education for internal medicine resident physicians specifically, based on the core competencies developed as outlined above. We provide an outline to assist in the actual implementation of such a curriculum within an internal medicine residency training program.

We have organized the curriculum into modules, with didactic as well as experiential components, utilizing a variety of educational venues, some new and some typically found within the existing framework of an internal medicine residency curriculum. Residency training programs may opt to deliver this curriculum via a dedicated rotation. While a dedicated rotation may be more efficient, it may also be less likely to be implemented as required components of residencies already replete with such rotations (e.g., intensive care unit). More importantly, we believe that addiction medicine is best taught to medical residents when the training is *integrated* into general medical care, modeling comprehensive care delivery. In this format, components of the unhealthy substance use curriculum are inserted into existing internal medicine teaching venues, both didactic and clinical, and the competencies contribute to the core general competencies addressed by residencies and monitored by the ACGME.

This model relies heavily on faculty who are well trained in addiction medicine and can serve as effective teachers. Such a model may also provide only limited exposure to patients in recovery after having received specialty treatment, who are less often recognized in internal medicine clinical settings. Finally, we link the proposed curricular modules to the ACGME core competencies. These modules may be modified and adapted to meet specific program needs and available resources.

Goal of the unhealthy substance use curriculum The goal of an unhealthy substance use (SU) curriculum for internal medicine residents is two-fold. The first goal is to highlight the importance of addiction medicine in patient care. The second is to provide internal medicine residents, regardless of ultimate career choice, with the core knowledge and skills necessary for all internists who provide clinical care. Of note, internists include those in general internal medicine (many of whom deliver primary medical care) as well as subspecialists (e.g., cardiologists, gastroenterologists, endocrinologists, nephrologists). Unhealthy SU condition knowledge and skills address appropriate prevention, early detection, diagnosis, treatment and referral for patients with substance use conditions. We outline herein a curriculum in unhealthy SU for internal medicine residents, based on the recommendations of AMERSA's Project MAINSTREAM regarding physician competencies. These **core competencies in unhealthy substance use for internal medicine residents** are as follows:

1) Residents will perform age, gender and culturally appropriate unhealthy substance use screening
2) Residents will effectively assess patients with unhealthy substance use
3) Residents will provide brief interventions to patients with unhealthy substance use
4) Residents will demonstrate effective counseling methods to help prevent unhealthy substance use
5) Residents will refer patients with substance use disorders to treatment settings that provide pharmacotherapy for relapse prevention
6) Residents will recognize, treat or refer co-morbid medical and psychiatric conditions in patients with substance use conditions
7) Residents will refer patients with substance use disorders to appropriate treatment and supportive services
8) Residents will be aware of the ethical and legal issues around physician impairment from substance use and of resources for referring potential impaired colleagues, including employee assistance programs, hospital based committees, and state physician health programs and licensure boards

9) Residents will identify the legal and ethical issues involved in the care of patients with unhealthy substance use
10) Residents will provide pharmacologic withdrawal to patients with substance dependence
11) Residents will provide or refer for treatment for relapse prevention in patients with substance use disorders, both pharmacotherapy and psychosocial counseling

Unhealthy substance use curriculum The curriculum is presented in eight modules, outlined in Table 1. The modules address the substance use competencies outlined by Project MAINSTREAM and are linked to the ACGME competencies that they can meet. Table 1

Table 1 Unhealthy substance use curricular modules, corresponding ACGME competencies and suggested internal medicine residency clinical venues

Modules	ACGME Competencies	Suggested Clinical Venue	Time
1) Addiction and the brain: principles of addiction	Medical knowledge; Patient care	All patient care activities, especially continuity clinic experiences; inpatient medicine; emergency department	45-60 minute lecture
2) Complications and comorbidities of unhealthy substance use	Medical knowledge; Patient care; Systems-based practice	Medical consultation rotations; inpatient medicine; emergency department; intensive care unit rotations	45-60 minute case-based lecture
3) Screening and assessment of unhealthy substance use	Practice-based learning;	Continuity clinic; inpatient medicine; subspecialty electives; emergency department	45-60 minute interactive lecture and 45-60 minute skills practice session
4) Effective methods of counseling patients including brief intervention	Patient care Medical knowledge; Interpersonal and communication skills; Systems-based practice	Continuity clinic; inpatient medicine; emergency department	45-60 minute interactive lecture and 45-90 minute skills practice (1-2 sessions)
5) Substance abuse treatment including pharmacotherapy	Systems-based practice; Medical knowledge; Patient care	Intensive care units, medical wards, continuity clinics, emergency department	60-90 minute lecture
6) Substance-specific inpatient and outpatient management	Medical knowledge; Patient care; Practice-based learning; Systems-based practice; Interpersonal and communication skills	Continuity clinic; inpatient medicine; emergency department	60 minute case-based lecture
7) Prescription drug abuse	Professionalism; Interpersonal and communication skills; System-based practice; Practice-based learning; Patient care; Medical knowledge	Continuity clinic; emergency department	60 minute case-based lecture
8) Legal and ethical considerations for patients and physicians	Professionalism; Systems-based practice; Practice-based learning; Medical knowledge	All patient care activities; continuity clinic; inpatient medicine; emergency department	60 minute lecture

suggests where to integrate the modules into an internal medicine residency program's existing clinical venues and an anticipated minimum time for each. The didactic sessions for each of the eight modules listed in Table 1 may be appropriate for more than one didactic session depending on the depth of teaching and the availability of group learning conference time. The curriculum is presented in separate modules, as a suggestion for implementation. However, these may be modified based on the needs of the individual training programs. For example, a residency program might opt to consolidate the screening and brief intervention modules into one didactic session, followed by a skill practice workshop.

1) Addiction and the brain: principles of addiction medicine

Module relevance This introductory module sets the stage for covering substance use conditions as important and relevant, often chronic conditions, for generalists. It will teach learners the pathophysiology of addiction.

Module content The session should introduce addiction as a chronic relapsing brain disease by reviewing the neurobiology of addiction. The neurobiology of addiction should include where in the brain substances of abuse act, how they cause intoxicating effects and how they alter the brain when used chronically. In this way, addiction medicine is framed similarly to how other chronic diseases (e.g., chronic obstructive lung disease, congestive heart failure, and diabetes) are taught. The session should also include national and local epidemiology of substances of abuse and describe the full spectrum of unhealthy substance use from at-risk use (that risks consequences) to substance dependence. A general overview about how the severity of the substance use problem influences treatment choice and efficacy and relapse risk should be covered. The session should be clinically relevant to the learner, for example, epidemiology and treatment should be presented in a way that is relevant to the specific residency program's needs and to the specific residents' rotation. For example, if the module is being taught to an inpatient team in an urban setting where heroin dependence is prevalent, then the prevalence, neurobiology and management of heroin dependence in hospitalized patients should be included. Discussion about the genetic vulnerabilities, risk and protective factors of addiction should also be covered.

Module special considerations Depending on the population served by the residency program's clinical sites, the curriculum should address the specific needs for special populations (adolescent, geriatric, racial/ethnic/cultural groups) and for specific substances of abuse.

2) Complications and comorbidities of substance abuse

Module relevance This module serves to highlight the impact SU has on other common medical and psychiatric diseases. The module is also important because of the high prevalence of medical and psychiatric comorbidities in patients with substance use conditions. As such it has particular relevance for internal medicine physicians focused on medical conditions, and common psychiatric conditions such as depression and anxiety. Addressing substance use allows more effective management of these conditions.

Module content It will be important for learners to understand how substance use causes or worsens other chronic diseases (e.g., cirrhosis, cardiomyopathy, depression) and has important interactions with treatments for other chronic diseases (e.g., anticoagulation therapy, sedatives for anxiety, opioids for chronic pain). This module highlights how common medical conditions (e.g., hypertension, insomnia) can be adversely affected by substance use. In addition, injection drug use and risky sexual behavior during substance use has been associated with conditions such as endocarditis, hepatitis B and C and HIV/AIDS. Topics such as cocaine associated chest pain or injection drug using patients with fever are useful contexts to present this topic.

3) Screening and assessment for unhealthy substance use

Module relevance This module has relevance for residents as screening and early intervention for unhealthy substance use are recommended practices for all adults [54]. Many internal medicine patients are unrecognized and once identified the problem can be addressed to prevent and manage substance use conditions. It also serves to teach specific skills on how to detect unhealthy (covering the spectrum from "at risk use" to "dependence") alcohol and drug use, using appropriate screening tools based on their validity, applicability and purpose.

Module content This module should cover how unhealthy substance use meets the criteria for widespread screening based on high quality evidence (high prevalence, significant consequences, valid screening tests, effective and safe treatments, early identification and treatment are preferable). The evidence behind effective formal screening methods (rationale, utility, operating characteristics) should be covered. Learners should practice specific techniques (single item screening tests, quantity and frequency, CAGE, AUDIT, DAST) [1,55] demonstrating age, gender and culturally appropriate unhealthy substance use screening skills. Learners should appreciate the limitations of biological markers (e.g., urine drug testing, blood mean corpuscular volume, gamma-glutamyl-transferase, carbohydrate-deficient transferrin). This module should address steps to be taken to assess patient's severity of substance use and readiness to change their use in patients who screen positive. Using the stages of change model, learners should be able to assess a patient's readiness to change. Assessment should include identifying substance use disorders (e.g., whether the patient has dependence, or

abuse, any consequences, or no consequences but excessive use). Teaching of assessment should also cover patient factors that increase the risk of any use, such as pregnancy or trying to conceive; medications contraindicated with substance use (e.g., warfarin); medical conditions that contraindicate alcohol or drug use (e.g., hepatitis); blackouts; failed attempts to cut down; family history of substance conditions; injuries related to substance use; medical conditions that may be caused by substance use (e.g., hypertension, trauma, anxiety, sleep disorders); and behavioral problems that can result from or be worsened by substance use (e.g., problems with work, school, or family).

4) Effective methods of counseling patients including brief interventions

Module relevance This module covers the effectiveness and skills development of counseling to help prevent the development of or progression of unhealthy substance use using formal psychological counseling and brief interventions. Brief counseling is one of the key skills in the recommended practice of screening and intervention. Motivational Interviewing and brief counseling are particularly important skills for managing patients with unhealthy substance use because many such patients do not recognize their condition, and when they do, they may not be ready to change. Brief counseling can facilitate change in this context. These skills also have relevance to internal medicine practice beyond addressing substance use, as they are useful for medication adherence, and behavior change counseling in general.

Module content Residents should learn stages of change (precontemplation, contemplation, determination, action, relapse and maintenance) and appropriate counseling strategies including patient advice and education about harms and risks. Residents should learn the skills of patient centered motivational interviewing and how they differ from confrontational approaches. They should be able to apply the principles of motivational interviewing including developing discrepancy, avoiding argumentation, rolling with resistance, expressing empathy and supporting self efficacy. They should be able to ask open ended questions, listen reflectively, affirm, summarize and elicit and recognize change talk (i.e., disadvantages of the status quo, advantages of change, optimism for change or intention to change; desire, ability and reasons for change statements, and commitment language). Residents should also be skilled in helping to strengthen a patient's commitment to change by negotiating a plan. Residents should learn the skills of brief intervention (i.e., counseling) including the components of patient feedback, emphasizing personal responsibility for change, giving clear advice, giving a menu of treatment options, having an empathic counseling style and enhancing a patients self-efficacy.

Module special considerations This module is best accomplished by employing skills practice where residents have a chance to role play brief intervention and motivational interviewing skills and receive feedback on their clinical skills.

5) Substance Abuse Treatment including Pharmacotherapy

Module relevance This module includes treatments that internal medicine physicians should have expertise in (e.g., pharmacotherapy for alcohol dependence) as well as treatments to which internists will generally refer patients (e.g., residential addiction specialty treatments).

Module content The module covers the effectiveness and content of substance abuse treatments including detoxification, (i.e., medically supervised withdrawal) residential treatment, 12 step and mutual help programs, outpatient treatment and pharmacotherapy (e.g., methadone, buprenorphine, naltrexone, acamprosate). Residents should learn how to work collaboratively with substance abuse specialty treatment programs and clinicians and specialists including counselors, psychologists and social workers. Emphasis should be on talking to patients about specialty treatment, making appropriate referrals and prescribing medications to treat dependence. Residents should know the efficacy and limitations of different treatment modalities. They should be familiar with web-based substance abuse treatment locator resources (e.g., Substance Abuse and Mental Health Services Administration (SAMHSA) homepage http://www.samhsa.gov).

Module special considerations Residents should have an opportunity to visit treatment programs (e.g., methadone maintenance program, alcoholics anonymous meeting) and interview patients who have undergone specialty treatment and/or attended mutual help groups, and those who are in recovery regardless of treatment history (many such patients can be found in the residents usual internal medicine training sites).

6) Substance Specific Inpatient and Outpatient Management

Module relevance This module serves to help residents identify specific substance intoxication and withdrawal syndromes, which are often seen in emergency, outpatient and inpatient medicine settings, and make evidence-based decisions on management strategies for specific substance (e.g., alcohol, opioid) intoxication, withdrawal and dependence.

Module content This module should cover substance specific epidemiology, biochemistry, clinical pharmacology (e.g., pharmacokinetics, drug testing, drug-drug interactions), neurobiology, and behavioral effects (e.g., intoxication, tolerance, physical dependence). The major substance categories should be covered including central nervous system depressants, psychomotor stimulants, nicotine, opioids, cannabis and alcohol. Residents should

be able to identify the signs and symptoms of intoxication, overdose and withdrawal of all the major categories of substances. Residents should learn how to manage acute intoxication and withdrawal syndromes of all the major categories of substances.

7) Prescription Drug Abuse

Module relevance The challenges of appropriate chronic pain treatment and recognition and prevention of prescription drug abuse are well known to internal medicine physicians. This module serves to give residents the knowledge and skills to prevent, and identify and manage a condition that is increasing in prevalence, prescription drug abuse (PDA).

Module content This module should include an overview of PDA including epidemiology, and important definitions (e.g., prescription drug misuse, tolerance, physical dependence, aberrant medication taking behavior, pseudoaddiction). Residents should know which medications are more likely to be abused and factors that lead to physicians overprescribing controlled substances. Residents should learn a framework for safe prescribing including understanding when controlled substances are indicated, and what their efficacy is, screening patients for PDA risk, setting realistic therapeutic goals and monitoring strategies including urine drug testing, pill counts, use of patient agreements and informed consent and use of prescription monitoring programs. Residents should be able to identify prescription drug abuse and have the skills to communicate with patients about either the lack of benefit from the controlled substance or apparent harm (e.g,, the possible diagnosis of addiction) to controlled substances. While not specifically part of a substance use curriculum, this module should be complemented by education on the treatment of acute and chronic pain.

8) Legal and Ethical Considerations for Patients and Physicians

Module relevance Documentation of substance use and care for substance use occurs in internal medicine patient encounters from screening for unhealthy SU to referral and treatment for dependence. Care for patients with unhealthy SU often involves challenges related to family and employment which can raise ethical issues. As physicians competent to recognize unhealthy substance use, internists are in a position to recognize it in their colleagues who may need help. Residents should be aware of the ethical and legal issues around caring for patients with substance use conditions as well as issues around physician impairment.

Module content This module should cover the patient confidentiality laws pertaining to managing patients with substance use conditions. Residents should also become aware of insurance coverage issues pertaining to substance use condition treatment and recognize that this coverage varies widely. Resources for information and updates on these topics should also be presented. Residents should be aware of the ethical and legal issues around physician impairment from substance use and of resources for referring potentially impaired colleagues including employee assistance programs, hospital-based committees, state physician health programs, and licensing boards.

Educational Venues

There are two components to the recommended curriculum-Didactic sessions and Clinical experiences.

Didactic sessions The core topics on substance use conditions outlined in each module may be presented using the teaching conferences already in place in internal medicine residency programs (e.g., departmental grand rounds, morbidity and mortality conference, noon conferences, etc.). In fact, doing so has the distinct advantage of treating unhealthy substance use in the same way that other medical conditions are treated in the residency curriculum. The didactic sessions will address concepts that will be reinforced during clinical rotations, or will provide a venue to address issues of addiction medicine that the residents may not be directly exposed to in their clinical rotations. Examples of common internal medicine residency didactic venues that can be used for unhealthy substance use teaching, and how they might be used, are:

- *Lectures and Morbidity and Mortality Rounds/ Conferences*: Overview of medical conditions with associated with substance abuse; overview of screening for substance abuse; detoxification procedures for alcohol and other drugs; medication management of addictions including relapse prevention; overview of appropriate prescribing practices for opioids and effective pain management strategies.
- *Grand Rounds*: As what is often the main academic conference within an entire department, this is an excellent venue for high profile, scientific presentations on unhealthy substance use.
- ***Case discussions****, such as Morning Report (inpatient and ambulatory), attending rounds or continuity clinic conferences*: Topics in unhealthy substance use may be presented *de novo* or previously presented topics can be reinforced during these case discussions. This is accomplished by discussing not only patients' medical conditions, but also by highlighting the underlying substance use conditions that may be associated with (having caused or worsened) them. Examples include: hepatitis C (injection drug use), cardiomyopathy (alcohol), and rhabdomyolisis (cocaine). During case presentations, any aspect may be highlighted along the spectrum of unhealthy SU, integrating substance use management principles seamlessly into the clinical discussions.

• *Journal Clubs*: Presentations would focus on critical appraisal skills while also addressing and reinforcing concepts in addiction medicine, through review of peer-reviewed articles on unhealthy substance use in the medical literature. Examples include: studies of the effectiveness of brief interventions or use of buprenorphine.

Additional teaching opportunities for addressing unhealthy substance use beyond traditional internal medicine residency conferences include:

• *Quality Improvement (QI) Projects*: With the implementation of an unhealthy substance use curriculum, residents may have the opportunity to meet the ACGME requirement of a QI project, within their own continuity clinic practice, or inpatient experience. Projects related to an assessment of institutional or administrative systems affecting implementation of screening, treatment protocols for substance withdrawal, or availability of treatment referrals for their patients are examples.
• ***Workshops***: Programs that include workshops or seminars (e.g., multi-hour small group sessions) for skills development can incorporate skills practice sessions in motivational interviewing and brief counseling interventions, as well as in the use of screening tools and approaches to assessment.
• ***Field Trips***: Visits outside the medical residency training program clinic and hospital setting to 12-step meetings, [56] methadone and buprenorphine clinics, needle exchange programs, other substance abuse treatment programs broaden the residents' view of the spectrum of substance use conditions and their treatment. These activities can be incorporated into conference series, ambulatory block experiences and seminars, inpatient rotations, and electives.
• ***Meetings with patients in recovery***: Residents can hear first-hand accounts of effective and ineffective ways that physicians approached their substance use condition from the patients themselves, creating powerful learning experiences. They can also be exposed to patients who are no longer severely affected. This venue may be incorporated into resident clinic conferences, including small or large group teaching sessions.
• *Video-taped and other observed patient encounters*: Direct observation of a patient encounter is the "gold standard" to assess residents' attitudes, knowledge and skills in the area of patient-doctor communication, and can teach and assess skills in screening, motivational interviewing and brief intervention. Guided by a trained preceptor, video-taped review is a powerful learning experience.

Clinical experiences Core concepts, presented in didactic sessions, can then be reinforced, and skills practiced, during residents' routine clinical rotations (inpatient medicine rotations, continuity clinics, emergency department, and intensive care unit rotations), complementing rather than replacing current curricular components, and delivered over the three years of training. Examples of experiential clinical learning experiences and opportunities include:

• *Inpatient hospital service rotations and intensive care rotations*: In these settings, residents can gain hands-on experience in the management of withdrawal syndromes (particularly opioids, alcohol and other sedatives), learn to assess the severity of addiction and readiness to change, and recognize and manage the complications and co-morbidities associated with substance use conditions, highlighting the importance of addressing the underlying substance use conditions. This is also a venue where resident physicians can have experiences and learning at the interface between pain and addiction.
• *Continuity clinics*: The resident continuity clinic is a site particularly well suited for the development of screening skills (and implementation–Practice-based learning), providing faculty with the opportunity to directly observe the resident selecting and performing the screening test. Specific curricula in patient-physician communication skills can easily incorporate a focus on screening for substance use conditions as one of the case examples. Screening and ongoing medical management of outpatients, referral to specialty treatment services and use of the electronic medical record-based screening and assessment systems are all learning opportunities that may be present at outpatient sites.
• *Emergency department*: Residents have the opportunity to assess patients with substance use conditions, and their consequences, both medical and social, and to identify existing local resources for referral to treatment. Residents will also see acute overdose and intoxication in this setting.
• *Medical consultation rotations/curriculum*: Residents may assist in the management of withdrawal syndromes on non-medical services, address perioperative issues of patients with substance use conditions, as well as the medical complications of addictions.
• *Specialty addiction treatment experiences*: We recommend that all residents experience at least one of the following specialty addiction treatment settings (to which they will refer patients, and from which they will receive patients), to broaden their view of addiction and recovery and to understand

the specialty care perspective. These programs include detoxification programs, needle exchange and methadone programs, 12-step meetings, residential rehabilitation programs and intensive outpatient treatment settings. These may be incorporated into an existing elective or ambulatory block, and need not be lengthy to be effective [56]. Residents' experience with substance use conditions would be further enriched by allowing residents to see patients in these settings, especially those in recovery who are doing well (who are often unlike patients residents recall from their internal medicine settings and experiences).

Evaluation of the Curriculum

The ability to demonstrate the achievement of competency-based learning objectives provides evidence that, when training is complete, graduating resident physicians can meet the health needs of the public [57]. In concordance with the ACGME's focus on *outcomes*, evaluation modalities should focus on whether and how fully internal medicine residents are incorporating concepts of addiction medicine into their practice, and whether residents are competently performing specific skills (e.g., screening and brief intervention) documented by direct observation and/or evidence in work products. Dimensions of evaluation of the unhealthy substance use curriculum include:

- Assessment of the effectiveness of didactic teaching
- Assessment of resident skill acquisition
- Clinical performance assessment and feedback
- Resident self-assessment and reflection
- Documentation of academic work products

Evaluation of the didactic sessions To measure learning in didactic presentations, before-after ("pre-post") measures can be administered with each conference to detect changes in targeted knowledge, skills (or perceived skills), and attitudes. Future intentions can also be measured to determine if teaching influenced residents' plans for changes or improvements in clinical practice. However, sometimes pre- post testing is difficult due both to time pressures and the common tendency for some residents to arrive to conferences late and others to leave early. Audience response technology, such as TurningPoint® (http://www.turningtechnologies.com accessed October 19, 2008), could be used as an alternative thereby integrating testing into an interactive presentation. This method would allow for the collection of real-time data using multiple choice questions near the beginning and the end of a teaching session. The quality and usefulness of the teaching can also be evaluated at the end of the didactic session employing a short resident questionnaire. Such evaluations are best kept to 3-5 questions to improve response rate.

Direct observation of resident skills A "mini-CEX (Clinical Evaluation Exercise)"-style evaluation card can be designed specifically for observation and evaluation of key addiction medicine clinical skills. The mini-CEX is a snapshot of doctor/patient interaction, designed to assess the clinical skills, attitudes and behaviors of trainees essential to providing high quality care by supervisors observing an actual clinical encounter. Not all elements need be assessed on each occasion. Specific unhealthy substance use clinical skills can easily be incorporated into routinely performed CEXs. The mini-CEX approach to resident clinical skills assessment is a feasible, reliable, valid, and widely used evaluation method [58]. For unhealthy substance use patients, the mini-CEX would be customized to focus on key substance use condition interactions and patient cases and different evaluation cards could be customized for both the inpatient and ambulatory setting.

If the appropriate equipment is available in the clinic setting and patient consent is obtained, video recording of encounters where screening or other targeted addiction medicine skills are employed provides the resident with the opportunity to self-assess and provides the faculty preceptor the opportunity to offer detailed feedback on resident performance. The key is to set aside sufficient time with resident and preceptor to review the video interaction in detail, stop and start the interaction frequently for reflection and skill assessment, and construct a dialogue focusing on reinforcement of skillful performance and opportunities for performance improvement.

Performance measures Teaching clinics may also opt to collect data on total number of patients screened and other individual resident performance statistics (e.g., brief intervention, referral to treatment). Data can be collected via medical record review and fed back to the resident. Key record indicators that can be measured include documentation of screening results, documentation of brief intervention provided, referrals, and follow-up.

Peer record reviews Peer medical record review can be a successful teaching strategy. Working in dyads, each resident conducts an annual review of a record of a patient with unhealthy substance use (or a patient screened for the condition)[59]. Residents select the record for their colleague to review, and together, they prepare a presentation for a pre-clinic conference, identifying challenges, treatment options, and community resources. In the process, residents have opportunity for reflection, self-assessment, and specific feedback on their approach to care and/or management.

Objective Structured Clinical Exams (OSCEs) The OSCE is an evaluation methodology where standardized patients (actors playing patient roles) are placed in mock healthcare settings to assess residents' clinical

skills. Standardized patients are trained to portray specific patient case scenarios in a standardized manner [60] so that clinical challenges can be consistently presented to the participating residents. OSCEs using standardized patients are employed for clinical skills assessment in many US residency programs and most US medical schools [61-63]. For these exams, residents engage in a clinical interaction with the standardized patient that is observed and/or video/audio-recorded for skill assessment by faculty. Standardized patients can also be trained to provide reliable assessments of residents' performance. The ACGME has identified OSCEs as the most desirable method for assessing interviewing, communication and counseling skills, and preventive health procedures, [64] and therefore can be an effective way to assess specific clinical skills key to substance use condition-related patient care. OSCEs are most cost-effective when a large number of residents participate during the same administration [64]. Resources needed for effective OSCE administration are substantial and include: space (a clinical facility or OSCE center), cases designed to call for the use of targeted unhealthy substance use skills, checklists for faculty and/or patients to use to assess targeted skills, trained standardized patients, and sufficient skilled faculty for observation or review of recorded interactions.

Portfolio submissions Portfolios are employed in resident evaluation for the documentation of clinical performance to meet competency criteria, documentation of program-related competence development, and documentation and tools for professional growth [65]. They are also commonly used to provide in-depth personalized mentoring to residents. There are a variety of different kinds of portfolios currently in use in residency training. Although many portfolios are paper, digital portfolios have also been shown to be well received in medical school and other settings, [66,67] as well as in residency training [68-71]. Digital portfolios can be employed not only to collect evidence of proficiency and professional growth, but can also enhance access and portability, organization, collaboration, and feedback [72].

For documentation and evaluation of addiction medicine skills, video recorded patient encounters on digital video discs (DVDs) or in digital format can be collected and maintained in the resident's academic portfolio. Pre-clinic conference presentations (slides, handouts, and presentation notes) can also be included. De-identified substance use condition management plans and reflective writings about related patient experiences or observations at AA meetings can also be valuable additions to resident portfolios. Portfolios are typically examined by program directors for the resident's annual review, are used as a basis for feedback on performance and resident professional development, and can be an important source of information for resident letters of recommendation for fellowship training or transition to private practice upon graduation.

Resources and implementation

Resources The delivery of the core curriculum, through our proposed modules would consist of, at minimum, 12 hours of didactics, over three years of residency training. Residents' same clinical experiences during standard internal medicine rotations would provide the clinical reinforcement of skills, and solidify the key concepts of addiction medicine/unhealthy substance use. Additional focused evaluation activities will reinforce both didactic and clinical teaching and provide outcome measures required by the ACGME. This time commitment represents only a fraction of the focus received by cardiology and diabetes care.

Costs Once a generation of internal medicine residency training program faculty with expertise sufficient to teach addiction medicine exist, teaching the required competencies to address unhealthy substance use should not require funds beyond those required for the whole program. Residency training programs often benefit from in-kind support or external support for special teaching efforts. Such added support would be useful for supporting experiences such as travel to treatment programs or standardized patients. However, these are the sorts of costs that, as for teaching of other competencies in other areas of medicine, are anticipated to be integrated with general funds for residency training, as part of the whole training program budget.

Achieving an adequate supply of internal medicine faculty capable of teaching unhealthy substance use competencies will be a cost. Unlike internal medicine subspecialty faculty, who are trained in medicine subspecialty fellowships and well-represented among internal medicine residency program faculty, medicine faculty with expertise in addiction medicine are not similarly well-represented. Such faculty will either need to come from new residency graduates, or from training current faculty, and in fairly large numbers. This training could occur by self-learning or specific continuing education experiences for existing faculty, or in national multidisciplinary training programs as have existed in the past and have been proposed recently (e.g., Centers of Excellence) [53].

Resources for Program Directors For training existing faculty, residency programs could link with local addiction treatment programs, and the few internal medicine or other faculty with this expertise in academic medical centers. Continuing medical education programs in person (e.g., as listed at http://www.motivationalinterview.org/ for motivational interviewing (accessed February 14, 2010), and perhaps more efficiently, online, exist for

training faculty. For example, the Alcohol Clinical Training Project (http://www.mdalcoholtraining.org accessed February 14, 2010) is a free resource for training faculty consisting of slides with case-based video vignettes, speaker notes and learner evaluation materials. The curriculum is flexible and modifiable and can be taught using all the components together in a 3-hour workshop or by using various components separately in 45 minute sessions (i.e., preclinic conference or attending rounds) or for self learning. The curriculum has been evaluated [37]. A related web publication ("Alcohol, other drugs and health: current evidence) can be used to for faculty (and residents) to keep up to date (http://www.aodhealth.org accessed February 14, 2008). Many teaching materials and information resources are available at http://www.nida.nih.gov, http://www.niaaa.nih.gov, and http://www.samhsa.gov (all accessed February 14, 2010). The NIAAA web site offers a clinician's guide to Helping Patients Who Drink Too Much and related continuing education materials (including slides and training videotapes).

Number of faculty needed Residency programs will need to have sufficient faculty (depending on the number of residents in the program) to assure their residents receive training, both to deliver the didactics, but even more importantly to mentor residents and serve as role models. Faculty serving as preceptors may need to be trained in the knowledge and skills slated for this curriculum. Thus faculty development efforts must be implemented, particularly when employing an integrated model. Residents value skill and competence in their teachers, and multifaceted teachers - those who have an excellent grasp on medicine, will be more effective teachers about unhealthy substance use for their residents.

Expertise Faculty teaching about unhealthy alcohol and drug use will need to have the competencies they are teaching (as noted previously). These faculty need not have addiction specialty expertise, and the expertise needed to address all competencies may be spread over a number of individuals. For example, one faculty member may have expertise in the management of inpatient alcohol withdrawal, another may be expert at screening and brief intervention. Many of the relevant competencies are similar to those that medicine faculty already must have to teach preventive medicine. These similar competencies are those that involve an understanding of modifiable risk factors and behaviors that risk chronic illnesses that are addressed by behavioral intervention aimed at lifestyle change (such as identification and addressing of depressive symptoms, medication nonadherence, physical inactivity). Such competencies are those reflected by ACGME competencies as communication skills (e.g., motivational counseling, for example is needed to address unhealthy drug use as well as lifestyle change and medication adherence). More teachers are needed to address competencies that all internists should have (e.g., prevention, screening, brief intervention, recognition of comorbidity, and ability to refer for pharmacotherapy and specialty care), whereas fewer teachers would be needed to address competencies needed to provide more specialized services such as prescription of buprenorphine for opioid dependence.

Required qualifications Recently, the American Board of Addiction Medicine was established to examine and certify diplomats (http://www.abam.net/ accessed February 24, 2010). While not yet recognized by the American Board of Medical Subspecialties (ABMS) it is the only US medical specialty board that certifies addiction medicine physicians across a range of medical specialties. The ABMS does recognize the specialty of addiction psychiatry (certification available to neurologists and psychiatrists) and the American Osteopathic Boards of Anesthesiology, Family Medicine, Internal Medicine and Neurology and Psychiatry recognize added qualifications in addiction medicine for their diplomates. Training and certification but not board certification can be obtained in other ways by (doctor of medicine) internists. In addition, much of the knowledge relevant to internists–the management of patients with unhealthy alcohol and other drug use in general populations and health settings–is not traditionally the focus of addiction specialty training (of note many leaders in the field of education and research on screening and brief intervention in the past 40 years have been generalist physicians).

The first qualification for training internal medicine residents is that the teachers be internal medicine physicians. It is important for internists to teach this (in distinction to psychiatrists or non-physicians) for several reasons. First, resident physicians see their mentors and attending physicians modeling appropriate care of patients with alcohol and drug problems. Second, internists are the most appropriate teachers to tailor the broad knowledge and evidence related to these conditions to the content most relevant for medicine residents and to the teaching venues most appropriate. Lastly, medicine teachers are most likely to demonstrate role responsibility for these conditions (and, as such, residency programs will take such responsibility). Interdisciplinary care and teaching is critical in the area of substance use conditions and faculty from other disciplines should be involved. Residents will most likely be receptive to such teaching when introduced by their primary clinical teachers (as other specialty subjects are taught during their residencies). There are parallels for this model in other areas of internal medicine such as depression and diabetes care.

Specialty credentials are not required for medicine faculty teaching about unhealthy SU. However,

continuing education courses and certification offered by the American Society of Addiction Medicine (ASAM) and attendance at professional continuing education meetings such as that of the Association for Medical Education and Research in Substance Abuse (AMERSA) can be helpful resources. Physicians teaching buprenorphine treatment of opioid dependence should be waivered by the Drug Enforcement Agency (DEA) by completing the required 8 hours of training or the other routes to achieve this goal (http://buprenorphine.samhsa.gov/index.html, accessed February 25, 2010). There have been numerous internists who have led teaching and educational efforts in internal medicine nationwide who have been successful without addiction specialty credentials or training. These teachers gain their expertise via generalist fellowships or focus on substance use during their residency training or via self- (by reading, online materials and workshops and courses at continuing education venues) and mentored teaching in their years as faculty.

Challenges in implementing new curricula in residency programs Internal medicine educators are bombarded with teaching requirements, for example in genetics, ethics, communication skills, molecular medicine, geriatrics, sexual health, computer literacy, and interdisciplinary team based care, to mention a few. These demands reflect continued rapid growth in scientific knowledge coupled with society's expectation that physicians minister to social and psychological as well as physical infirmities[73]. Training programs are challenged to produce competent, ethical and caring physicians in a defined time period of three years, but must address key health care needs of the population, among which unhealthy substance use certainly figures highly. As internal medicine educators, we cannot afford to ignore addiction medicine with the significant national burden of disease as well as proven treatment interventions. There is still a serious mismatch between what we know to be good quality care and the care that is actually delivered, [52] and residency programs are in a position to fill this gap. One way to address the challenge of a bursting curricula in residency training, is to integrate a new focus of discussion into standard teaching venues.

Champions For any curriculum change to be implemented and maintained there needs to be a motivated and vigilant faculty champion who is dedicated to providing appropriate intervention and support to the residency program as needed. Ideally, this champion would be a faculty member who is active in curriculum management and reform and willing to take on a leadership role. Designing an unhealthy substance use curriculum implementation strategy for any residency program will require a fundamental appreciation of the culture and resources of the program, its faculty, and its teaching venues, as well as an understanding of the addiction field itself. The champion would probably be active in teaching several didactic sessions to both residents and faculty as needed, and would monitor the status of the curriculum and its evaluation to fine-tune and improve the program over time.

Summary

Given the burden of disease and the effective interventions that can be delivered to patients by internal medicine physicians, teaching about unhealthy substance use must, and can be incorporated into internal medicine residency training. In part because clinical management for substance use conditions is most effective when integrated with medical and other care, education should mirror this approach, integrating curricula on unhealthy substance use into overall residency training. This integration is not particularly resource intensive and can be done by making use of many existing teaching venues. Perhaps the biggest challenge will be initially getting sufficient trained internal medicine faculty with expertise to teach the curriculum for residents. Getting sufficient faculty will likely require resources since unlike other medicine topics, a new faculty (or new expertise, or at least expertise that is not in abundant supply in internal medicine residency programs) is required. Many educational resources already exist to achieve this goal though the manpower and time for training will still require investment. Federally supported faculty development efforts have trained a small number of expert teachers. Programs similar to these will likely be needed to implement curricula such as we propose. Further, the development of addiction medicine as a specialty by the American Board of Addiction Medicine is likely to support implementation of such curricula in the future. With adequate faculty, few other challenges remain. Core unhealthy substance use competencies for physicians, easily adapted to internists, have been established, and they can contribute to meeting ACGME general competencies. A curriculum has been outlined herein, based on this guidance. Evaluation of outcomes is fairly straightforward and achievable. Accrediting bodies can have a significant role in improving the teaching and therefore care for patients with unhealthy substance use. The time is right to improve (and require excellence in) residency training about unhealthy substance use in internal medicine residency training programs.

Acknowledgements

This manuscript was written for the Betty Ford Institute Consensus Conference, December 2008. The Betty Ford Institute provided financial support for publication fees and for experts (including RS)to attend the Consensus Conference.

Author details
[1]Residency Training Program in Internal Medicine, Boston University School of Medicine, Boston, MA, USA. [2]Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, Boston University School of Medicine, Boston, MA, USA. [3]Department of Medicine, Boston Medical Center, and Boston University School of Medicine, Boston, MA, USA. [4]Youth Alcohol Prevention Center, Boston University School of Public Health, Boston, MA, USA. [5]Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA.

Authors' contributions
AHJ, DPA and RS contributed equally to this work, CED prepared the Evaluation of the Curriculum section. All authors were involved in revising the manuscript, and all authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Received: 9 June 2009 Accepted: 15 March 2010
Published: 15 March 2010

References

1. Saitz R: **Unhealthy alcohol use.** *N Engl J Med* 2005, **352**:596-607.
2. Cherry DK, Woodwell DA, Rechtsteiner EA: **National Ambulatory Medical Care Survey: 2005 Summary. Advance data from vital and health statistics.** *Hyattsville, MD* 2007, **387.**
3. SAMHSA: **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. Rockville, MD* 2007.
4. Office of National Drug Control Policy: **National Drug Control Strategy. Data Supplement.** *Washington, D.C* 2005 [http://www.whitehousedrugpolicy.gov/publications/policy/ndcs09/ndcs09_data_supl/ds_drg_rltd_tbls.pdf], Accessed October 19, 2008.
5. Mokdad AH, Marks JS, Stroup DF, Gerberding JL: **Actual causes of death in the United States, 2000.** *JAMA* 2004, **291**:1238-1245.
6. NIAAA: 2008 [http://www.niaaa.nih.gov/Resources/DatabaseResources/QuickFacts/EconomicData/cost8.htm], Access year August 25, 2008.
7. Greenlund KL, Giles WH, Keenan NL, *et al*: **Heart disease and stroke mortality in the 20th century.** *Silent Victories: The History and Practice of Public Health in Twentieth Century America* Oxford, England.: Oxford University PressWard J, Warren C 2006.
8. Rosamond W, Flegal K, Friday G, *et al*: **Heart disease and stroke statistics – 2007 update: a report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee.** *Circulation* 2006, **113**:169-171.
9. Fiellin D, Butler R, D'Onofrio G, Brown R, O'Connor P: **The physician's role in caring for patients with substance use disorders: Implications for medical education and training.** *Subst Abuse* 2002, **23**:207-212.
10. Michaud CM, Murray CJ, Bloom BR: **Burden of disease – implications for future research.** *JAMA* 2001, **285**:535-539.
11. Office of National Drug Control Policy: **The economic cost of drug abuse in the United States, 1992-2002.** *Washington, D.C* 2002 [http://www.ncjrs.gov/ondcppubs/publications/pdf/economic_costs.pdf], Accessed August 25, 2008.
12. SAMHSA: **Results from the 2006 national survey of drug use and health: national findings.** [http://www.drugabusestatistics.samhsa.gov/NSDUH/2k6NSDUH/2k6results.cfm], Accessed August 25, 2008.
13. Friedmann PD, Saitz R, Samet JH: **Management of adults recovering from alcohol or other drug problems: relapse prevention in primary care.** *JAMA* 1998, **279**:1227-1231.
14. Lewis DC: **Despite benefit, physicians slow to offer brief advice on harmful alcohol use.** *JAMA* 1996, **299**:751-753.
15. 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**:1689-1695.
16. Saitz R, Horton NJ, Larson MJ, Winter M, Samet JH: **Primary medical care and reductions in addiction severity: a prospective cohort study.** *Addiction* 2005, **100**:70-78.
17. Samet JH, Friedmann PD, Saitz R: **Benefits of linking primary medical care and substance abuse services: patient, provider, and societal perspectives.** *Arch Intern Med* 2001, **161**:85-91.
18. 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)**:143-152.
19. Maciosek MV, Coffield AB, Edwards NM, Flottemesch TJ, Goodman MJ, Solberg LI: **Priorities among effective clinical preventive services: results of a systematic review and analysis.** *Am J Prev Med* 2006, **31(1)**:52-61.
20. Maciosek MV, Coffield AB, Edwards NM, Goodman MJ, Flottenmesch TJ, Solberg LL: **Priorities among effective clinical preventive services: results of a systematic review and analysis.** *Am J Prev Med* 2006, **31(1)**:52-61.
21. McCormick KA, Cochran NE, Back AL, Merrill JO, Williams EC, Bradley KA: **How primary care providers talk to patients about alcohol: a qualitative study.** *J Gen Intern Med* 2006, **21**:966-972.
22. Kuehn BM: **Despite benefit, physicians slow to offer brief advice on harmful alcohol use.** *JAMA* 2008, **299**:751-753.
23. Vastag B: **Addiction poorly understood by clinicians: experts say attitudes, lack of knowledge hinder treatment.** *JAMA* 2003, **290**:1299-1303.
24. Saitz R, Friedmann PD, Sullivan LM, *et al*: **Professional satisfaction experienced when caring for substance-abusing patients: faculty and resident physician perspectives.** *J Gen Intern Med* 2002, **17**:373-376.
25. Stimmel B, Cohen D, Swartz M: **An assessment of house staff's knowledge of alcohol and substance abuse utilizing standardized patients.** *Subst Abuse* 2000, **21**:1-7.
26. Institute of Medicine: **Improving the quality of health care for mental and substance-use conditions.** Washington, DC, The National Academics Press 2006.
27. Conigliaro J, Delos Reyes C, Parran TV, Schulz JE: **Principles of screening and early intervention.** *Principles of Addiction Medicine* Chevy Chase: American Society of Addiction MedicineGraham AW, Schultz TK, Mayo-Smith MF, Ries RK, Wilford BB , Third 2003.
28. Office of National Drug Control Policy: **Leadership Conference on Medical Education in Substance Abuse December 1-2, 2004 (report).** *Washington, D.C* 2004 [http://www.ncjrs.gov/ondcppubs/publications/pdf/medical_educ_2004.pdf], Access date, October 19.
29. Fleming MF, Barry KL, Davis A, Kropp S, Kahn R, Rivo M: **Medical education about substance abuse: changes in curriculum and faculty between 1976 and 1992.** *Acad Med* 1994, **69**:362-369.
30. Fleming MF, Manwell LB, Kraus M, Isaacson JH, Kahn R, Stauffacher EA: **Who teaches residents about the prevention and treatment of substance use disorders? A national survey.** *J Fam Pract* 1999, **48**:725-729.
31. Foley ME, Garland E, Stimmel B, Merino R: **Innovative clinical addiction research training track in preventive medicine.** *Subst Abuse* 2000, **21**:111-119.
32. Kuehn BM: **Centers to weave addiction treatment into medical education.** *JAMA* 2007, **297**:1763.
33. 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**:410-418.
34. Wyatt SA, Vilensky W, Manlandro JJ Jr, Dekker MA: **Medical education in substance abuse: from student to practicing osteopathic physician.** *J Am Osteopath Assoc* 2005, **105**:S18-S25.
35. Klamen DL, Miller NS: **Integration in education for addiction medicine.** *J Psychoactive Drugs* 1997, **29**:263-268.
36. Project Mainstream: **Improving substance abuse education for health professionals.** [http://www.projectmainstream.net], Accessed March 11, 2008.
37. Alford DP, Richardson JM, Chapman SE, Dube CE, Schadt RW, Saitz R: **A web-based alcohol clinical training (ACT) curriculum: Is in-person faculty development necessary to affect teaching?** *BMC Medical Education* 2008, **8**:11.
38. Parish SJ, Ramaswamy M, Stein MR, Sachur EK, Arnsten JH: **Teaching about substance abuse with objective structured clinical exams.** *J Gen Intern Med* 2006, **21**:453-459.
39. Alford DP, Compton P, Samet JH: **Acute pain management for patients receiving maintenance methadone or buprenorphine therapy.** *Ann Intern Med* 2006, **144**:127-134.
40. Samet JH, Rollnick S, Barnes H: **Beyond CAGE. A brief clinical approach after detection of substance abuse.** *Arch Intern Med* 1996, **156**:2287-2293.
41. Abrams Weintraub T, Saitz R, Samet JH: **Education of preventive medicine residents: alcohol, tobacco, and other drug abuse.** *Am J Prev Med* 2003, **24(1)**:101-105.

42. Isaacson JH, Fleming MF, Kraus M, Kahn R, Mundt M: **A national survey of training in substance use disorders in residency programs.** *J Stud Alcohol* 2000, **61**:912-915.
43. Chossis I, Lane C, Gache P, Michaud PA, Pecoud A, Rollnick S, Daeppen JB: **Effect of training on primary care residents' performance in brief alcohol intervention: a randomized controlled trial.** *J Gen Intern Med* 2007, **22(8)**:1144-1149.
44. Karam-Hage M, Nerenberg L, Brower KJ: **Modifying residents' professional attitudes about substance abuse treatment and training?** *Am J Addict* 2001, **10**:40-47.
45. Geller G, Levine DM, Mamon JA, Moore RD, Bone LR, Stokes EJ: **Knowledge, attitudes, and reported practices of medical students and house staff regarding the diagnosis and treatment of alcoholism.** *JAMA* 1998, **261**:3115-3120.
46. Madden TE, Graham AV, Straussner SL, *et al*: **Interdisciplinary benefits in Project MAINSTREAM: a promising health professions educational model to address global substance abuse.** *J Interprof Care* 2006, **20**:655-664.
47. Marcus MT, Brown RL, Straussner SL: **Creating change agents: a national substance abuse education project.** *Subst Abuse* 2005, **26**:5-15.
48. 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)**:280-295.
49. Physician Quality Reporting Initiative: **Centers for Medicare and Medicaid Services.** [http://www.cms.hhs.gov/pqri], Accessed October 28, 2009.
50. **Candidate Measure Profile for Assessing and Treating Tobacco, Alcohol, and Other Drug Use and Dependence.** [http://www.jointcommission.org/NR/rdonlyres/3528C77F-7BB5-4D62-85E9-87C8285D8652/0/CandidateFinalMIFTADD6.doc], accessed September 29, 2009.
51. Yoast RA, Wilford BB, Hayashi SW: **Encouraging physicians to screen for and intervene in substance use disorders: Obstacles and strategies for change.** *J Addict Dis* 2008, **27(3)**:1299-1303.
52. Greiner AC, Knebel E, (Eds): *Committee on the Health Professions Education Summit* Washington, D.C.: National Academics Press 2003.
53. Haack ML, Adger HA, (Eds): *Strategic Plan for Interdisciplinary Faculty Development; Arming the Nation's Health Professional Workforce for a New Approach to Substance Use Disorders* 2002, **23.**
54. Whitlock EP, Polen MR, Green CA, Orleans T, Klein J: **Behavioral counseling interventions in primary care to reduce risky/harmful alcohol use by adults: a summary of the evidence for the U.S. Preventive Services Task Force.** *Ann Intern Med* 2004, **140(7)**:557-568.
55. Skinner HA: **The drug abuse screening test.** *Addict Behav* 1982, **7(4)**:363-367.
56. Rose AJ, Stein MR, Arnsten JH, Saitz R: **Teaching internal medicine resident physicians about alcoholics anonymous: A pilot study of an educational intervention.** *Subst Abuse* 2006, **26(3)**:5-11.
57. Lynch DC, Swing SR: **Research Department ACGME.** *Key Considerations for selecting assessment instruments and implementing assessment systems* 2008 [http://www.acgme.org/outcome/assess/keyConsider.asp], accessed October 29, 2008.
58. Norcini JJ, Blank LL, Duffy DF, Fortna GS: **The Mini-CEX: a method for assessing clinical skills.** *Ann Intern Med* 2003, **138**:476-481.
59. Gillig PM, Barr A: **A model for multidisciplinary peer review and supervision of behavioral health clinicians.** *Community Mental Health J* 1999, **35(4)**:361-365.
60. Stimmel B, (Ed): *Utilizing standardized patient protocols to improve skills in identifying tobacco, alcohol and other drug use: A manual of cases* New York: Josiah Macy Jr. Foundation 1998.
61. ACGME/ABMS Joint Initiative: **Suggested best methods for evaluation. Attachment/Toolbox of Assessment Methods. version 1.1.** 2000 [http://www.acgme.org/outcome/assess/toolbox.asp], Accessed 10/16/08.
62. Stillman PL, Swanson DB, Regan MB, *et al*: **Assessment of clinical skills of residents utilizing standardized patients. A follow-up study and recommendations for application.** *Ann Intern Med* 1991, **114(5)**:393-401.
63. Stillman PL, Swanson DB, Smee S, *et al*: **Assessing clinical skills of residents with standardized patients.** *Ann Intern Med* 1986, **105(5)**:762-771.
64. ACGME Outcomes Project, American Board of Medical Specialties: **Toolbox of assessment methods: A product of the joint initiative.** 2000 [http://www.acgme.org/Outcome/assess/ToolTable.pdf], Accessed 10/16/08.
65. Tillema H: **Portfolios as developmental assessment tools.** *Int J Training Dev* 2001, **5(2)**:126-135.
66. Dreissen EW, Muijtjens AM, van Tartwijk J, Vleuten van der CP: **Web- or paper-based portfolios: is there a difference?** *Med Educ* 2007, **41**:1067-1073.
67. Woodward H, Nanlohy P: **Digital portfolios: fact or fashion?** *Assess Eval Higher Educ* 2004, **29(2)**:227-238.
68. Carraccio C, Englander R: **Evaluating competence using a portfolio: a literature review and web-based application to the ACGME competencies.** *Teaching Learning Med* 2004, **16(4)**:381-387.
69. Clay AS, Petrusa E, Harker M, Andolsek K: **Development of a web-based, specialty specific portfolio.** *Med Teacher* 2007, **29**:311-316.
70. Fung MFKF, Walker M, Fung KFK, *et al*: **An internet-based learning portfolio in resident education: The KOALATM multi-centre programme.** *Med Educ* 2000, **34**:474-479.
71. Rosenberg ME, Watson K, Paul J, Miller W, Harris I, Valdivia TD: **Development and implementation of a web-based evaluation system for an internal medicine residency program.** *Acad Med* 2001, **6**:92-95.
72. May K, Lowry B: **Digital Portfolios to Support Professional Development and Program Evaluation.** 2007.
73. Batalden P, Leach D, Swing S, Dreyfus H, Dreyfus S: **General competencies and accreditation in graduate medical education.** *Health Affairs* 2002, **21(5)**:103-111.

Pre-publication history
The pre-publication history for this paper can be accessed here:
[http://www.biomedcentral.com/1472-6920/10/22/prepub]

doi:10.1186/1472-6920-10-22
Cite this article as: Jackson *et al.*: Internal medicine residency training for unhealthy alcohol and other drug use: recommendations for curriculum design. *BMC Medical Education* 2010 **10**:22.

Original Paper

Caries Research

Caries Res 2010;44:421–430
DOI: 10.1159/000318569

Received: October 12, 2009
Accepted after revision: June 30, 2010
Published online: August 20, 2010

Snacking Habits and Caries in Young Children

I. Johansson[a] P. Lif Holgerson[a] N.R. Kressin[b, c, d] M.E. Nunn[b] A.C. Tanner[e, f]

[a]Department of Odontology, Umeå University, Umeå, Sweden; [b]Northeast Center for Research to Evaluate and Eliminate Dental Disparities (CREEDD), Goldman School of Dental Medicine, [c]Healthcare Disparities Research Program, Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, [d]VA Boston Healthcare System, [e]Department of Molecular Genetics, Forsyth Institute, and [f]Department of Oral Medicine, Infection and Immunity, Harvard School of Dental Medicine, Harvard University, Boston, Mass., USA

Key Words
Breast-feeding · Caries · Children · Snacks

Abstract
Dental caries is caused by a combination of infection and diet. This disease, if left untreated, may lead to pain, and impair the quality of life, nutritional status and development of young children. The objective was to investigate the association between snacking and caries in a population at high risk of dental caries. American preschool children (n = 1,206) were recruited in the offices of paediatricians. Data on sociodemographic characteristics, oral hygiene, breast-feeding, use of bottle and snacking were collected by questionnaire. Plaque presence, the number of teeth and their caries status (deft) were scored. The children sampled were 61% Black, 27% White and 10% Asian. Of the 1- to 2-, 2- to 3- and 3- to 4-year-old children, 93.8, 82.4 and 77.3% were caries free, and their mean caries scores were 0.16, 0.58 and 0.93, respectively. Multivariate partial least squares (PLS) modelling revealed plaque presence, lowest income, descriptors for tooth exposure time (number of teeth and age) and cariogenic challenge (total intake of sugar-containing snacks and chips/crisps, and chips intake with a sugar-containing drink) to be associated with more caries. These differences were also found in univariate analyses; in addition, children who continued breast-feeding after falling asleep had significantly higher deft values than those who did not. PLS modelling revealed that eating chips clustered with eating many sweet snacks, candies, popcorn and ice cream. We conclude that, in addition to the traditional risk indicators for caries – presence of plaque, sugar intake and socioeconomic status –, consumption of chips was associated with caries in young children.

Dietary habits have shifted in all age groups in the Western populations in recent decades, including a nearly doubled intake of energy-dense, low-nutrient-dense snack foods [Briefel and Johnson, 2004; Adair and Popkin, 2005]. In children, more than 30% of the daily energy intake was reported to come from such foods and, on average, 75% of Americans report daily snacking [Briefel and Johnson, 2004]. In addition, different snacking patterns have been reported based on household income: individuals with income at or below the poverty line in the USA more frequently consumed po-

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

Accessible online at:
www.karger.com/cre

Ingegerd Johansson
Department of Odontology, Umeå University
SE–901 87 Umeå (Sweden)
Tel. +46 90 785 6035, Fax +46 90 770 570
E-Mail ingegerd.johansson@odont.umu.se

KARGER

tato chips, fried potatoes, whole milk and fruit drinks, whereas those with higher incomes consumed more grain-based salty snacks, fruits, skim milk, soft drinks, coffee and tea [Briefel and Johnson, 2004]. The shifted dietary patterns have been associated with increased risk of endemic diseases such as type 2 diabetes, obesity and dental caries.

Dental caries is a chronic infectious disease characterized by demineralization of tooth tissues at lowered pH following bacterial fermentation of dietary carbohydrates. The period of critically lowered pH needed for caries to occur is mainly a function of the type and frequency of carbohydrates consumed, the microbial composition of the tooth biofilm and salivary factors. Sucrose and monosaccharides induce a rapid and deep pH drop, and accordingly increase the risk of caries [Azevedo et al., 2005; Seow et al., 2009]. Energy-dense, low-nutrient-dense foods are often characterized by a high content of added sugar, but several modern snack products such as chips (crisps), popcorn and shrimp crackers, while not sweet, are still potentially cariogenic due to their content of extensively hydrolysed starch [Lingström et al., 2000]. Snacking has gained an increasing role as a risk indicator for caries development in children [Milgrom and Reisine, 2000; Marshall, 2005], but so far the impact of products with extensively hydrolysed starch, such as potato chips, on caries risk has not been demonstrated in children or adults.

Caries in young children, often referred to as early childhood caries [American Academy of Pediatric Dentistry, 2008], may, if left untreated, lead to pain, reduced quality of life and impaired eating, and may impair a child's nutritional status and development. The prevalence of early childhood caries varies between communities but is frequently high in underprivileged communities and among disadvantaged immigrants [Grindefjord et al., 1993; Milnes, 1996; Petersen and Esheng, 1998; Wennhall et al., 2002; Jose and King, 2003; Stecksen-Blicks et al., 2004; Vachirarojpisan et al., 2004; Beltrán-Aguilar et al., 2005; Schroth et al., 2005]. Understanding the role of lifestyle-associated risk indicators for dental caries in young children in groups prone to caries development forms the basis for targeted caries prevention programs. The aim of the present study has been to investigate the association between snacking habits and caries, considering oral hygiene and socioeconomics as possible confounders, in a cohort of young children living in a population at high risk of dental caries in an industrialized country (USA).

Subjects and Methods

Study Cohort

Preschool children presenting for well-child visits at the paediatric clinics at Boston Medical Center, Boston University, and the Floating Hospital, Tufts Medical Center, Boston, USA, were recruited in a 12-month period in 2003–2005. These hospitals serve all racial and ethnic groups living in the Boston area but have a special mission for underserved groups [Kressin et al., 2009]. Inclusion criteria were that a child was 6 months to no more than 5 years of age, and that a parent or guardian was willing to consent to the child's clinical examinations [Kressin et al., 2009]. Children with congenital diseases affecting the dentition were excluded. The study design, protocol, questionnaire and informed consent were approved by the institutional review boards of the institutions involved.

Data Collection

Data on sociodemographic characteristics (gender, family income, education level, race and ethnicity) and oral hygiene, feeding (breast-feeding or use of bottle) and snacking habits were obtained from parents or guardians, collected at the offices of paediatricians via a structured questionnaire [Kressin et al., 2009].

The number of teeth, and their status as sound, precavitated (white spot lesion), cavitated, filled or sealed was recorded [Drury et al., 1999], using good light, a disposable mirror and an explorer. For each child, the total number of decayed (d; non-cavitated and cavitated), filled (f; sealants not included) and extracted (e) primary teeth (t) was calculated. Visible plaque was recorded on a 0–3 scale (no plaque, and plaque covering a mean surface area of $<1/3$, 1/3 to $<2/3$ or $>2/3$ of the tooth) [Kanasi et al., 2010]. Data collection and oral examinations were performed by 2 specially research-trained dental hygienists.

Data Analyses

For descriptive data and associated univariate analyses, family income status was dichotomized as relatively high or low [equal to or higher than or below the median income in 2006 in the state of Massachusetts (USD 56,292 in 2008) according to the US Census Bureau news release (www.census.gov)], education to high (higher than high school) or low (equal to or lower than high school), and other variables as yes or no – for example, presence of caries (deft ≥ 1) or not (deft = 0), visible plaque (score ≥ 1) or no visible plaque (score 0), daily cleaning of teeth or not, and reporting eating a snack most days or not.

Categorical data are presented as proportions (percent children), and distribution differences between groups were tested by a χ^2 test with $p < 0.01$ considered statistically significant. Caries data (deft scores) are presented as means with 95% CI after standardization for age group or number of teeth in age-merged and age-stratified groups, respectively. Standardized least square means with 95% CI were calculated using the general linear model (GLM) procedure followed by the Bonferroni multiple mean test, with $p < 0.05$ considered statistically significant. The SPSS software (version 16.0.1 for Windows; SPSS Inc., Chicago, Ill., USA) was used.

Table 1. Description of study cohort

	1–2 years (n = 678)	2–3 years (n = 312)	3–4 years (n = 216)
Teeth	9.6 (9.3–10.0)	18.0 (17.8–18.3)	19.9 (19.8–20.0)
Caries[1]			
deft score	0.16 (0.11–0.21)	0.58 (0.39–0.78)	0.93 (0.58–1.28)
Caries free, %	93.8	82.4	77.3
Gender, %			
Boys	51.0	56.1	46.8
Girls	49.0	43.9	53.2
Race, %			
Black	58.1	61.2	70.8
White	28.5	28.2	22.2
Asian	12.5	10.3	6.9
Other	0.8	0.3	0
Ethnicity[2], %			
Hispanic	14.3	17.6	12.0
Non-Hispanic	85.7	82.4	88.0
Education, % ≤high school	45.8	48.7	54.0
Income[3], % low	71.3	72.8	75.0
Tooth cleaning[4], %			
Daily	73.2	92.6	96.8
Never	13.4	0.6	0.5
Visible plaque, % with score ≥1	23.0	49.8	44.2
Snack intake[5], % of children			
Fresh fruit	76.5	79.2	81.5
Crackers	69.9	68.6	65.7
Yoghurt	50.3	48.6	60.8
Cookies	47.8	47.4	60.2
Cereals (dry)	48.7	41.0	32.4
Chips	23.0	40.7	44.0
Cereals with milk	25.4	32.1	33.3
Ice cream	15.9	24.0	35.2
Candies	9.3	21.2	31.9
Dried fruit	11.2	17.3	15.7
Popcorn	5.3	15.4	27.6
Others	28.2	32.6	32.9
Breast-feeding, %			
Yes (currently)	8.9	3.7	1.7
Daily at sleep	7.3	3.0	1.2
Drink from bottle, %			
Yes (usually)	67.3	26.6	13.0
Bottle in bed	49.6	21.6	12.1

Values other than percentages denote means with 95% CI in parentheses.

[1] deft = sum of decayed, extracted (caries) and filled deciduous teeth; caries free = no detectable white or cavitated lesions, restorations or teeth extracted due to caries.

[2] <0.5% were native Hawaiian or American Indian, respectively.

[3] Low income = annual income less than median income in Massachusetts (USD 56,000).

[4] Remaining children cleaned teeth several times a week or month. Tested among all levels.

[5] Daily snacking was reported for 96, 97 and 99% of the children in the 3 age groups, respectively. Snacks reported to be eaten daily by a child are listed. The numbers indicate the proportion (%) of children for whom daily intake was reported (consumer).

Table 2. Caries status by gender, socioeconomic factors and oral hygiene measures

	Numbers	Children with caries[1]		Caries experience[2]	
		%	p	deft	p
Gender					
Boy	622	14.0	0.039	0.59 (0.47–0.71)	0.383
Girl	584	10.1		0.52 (0.38–0.66)	
Race					
Black	738	13.3	0.291	0.55 (0.43–0.67)	0.150
White	329	9.1		0.51 (0.33–0.69)	
Asian	132	13.6		0.73 (0.48–0.98)	
Other	7	0		0	
Ethnicity					
Hispanic	178	12.4	0.911	0.56 (0.32–0.80)	0.988
Non-Hispanic	1,028	12.1		0.56 (0.46–0.66)	
Education[3]					
Low	571	13.8	0.080	0.60 (0.46–0.74)	0.378
High	618	10.5		0.52 (0.38–0.66)	
Income[4]					
<Median (low)	704	14.5	0.001	0.68 (0.56–0.80)	0.009
≥Median (high)	269	7.1		0.38 (0.28–0.58)	
Plaque					
Yes	344	23.8	<0.0001	0.96 (0.84–1.08)	<0.0001
No	668	6.4		0.29 (0.13–0.45)	

[1] Crude numbers tested by χ^2 test for differences in group proportions. $p < 0.01$ for statistical significance.
[2] Means (95% CI) standardized for age group by the GLM procedure (standardization for number of teeth had no further effect). Differences between groups tested by the Bonferroni post hoc test. $p < 0.05$ for statistical significance.
[3] Low education defined as low when equal to or lower than high school.
[4] Low income defined as an annual income less than the median income in Massachusetts (USD 56,000).

Multivariate Analysis

Partial least squares (PLS) modelling using SIMCA P+ (v. 12.0; Umetrics AB, Umeå, Sweden) was used for multivariate analysis. PLS is a multivariate linear regression model method that detects correlations between matrices of independent and covarying descriptor and response variables. The variables used were snack items and risk indicators or factors for dental caries in small children. These were modelled using logarithmically (ln) transformed deft (after addition of 0.01 to all values) and the dichotomized caries score. All variables were autoscaled to unit variance.

The importance of each variable of interest (x-variable) in explaining the variation among the outcome variables (y-variables) is given by a correlation coefficient and a variable importance in projection (VIP) value. A VIP value of >1.0 is influential, and a VIP value of ≥1.5 highly influential. The R^2 and Q^2 values give the capacity of the x-variables to explain (R^2) and predict (Q^2) the variance among the y-value(s). Q^2 values, which preferentially should not differ by more than 0.2 from model R^2 values, were obtained by cross-validation where every 7th observation was kept out of the model and predicted by a model from the remaining observations. This was repeated until all observations had been kept out once.

Results

A total of 1,291 children were examined, but the data are restricted to 1,206 children because 13 children were predentate and 43 children non-cooperative (no information on tooth status); further dentate children younger than 1 year of age (n = 29) were omitted as the distribution of several variables deviated markedly from the other age groups, including lack of clinical caries measurements (data not shown). No child was older than 4 years.

Demographic and Oral Characteristics

The racial and ethnic distributions of the 1,206 children examined were 61% Black, 10% Asian, 27% White and 15% Hispanic (table 1). Boys and girls were equally represented. Approximately half of the parents or guardians had an education lower than or equal to high school, and more than 70% had an annual income below the me-

dian income in the state (table 1). The variations between the 3 age groups are shown in table 1.

The number of erupted teeth ranged from 1–20, with an average of 13.6 teeth (95% CI: 13.3–14.0) and an expected increase by age (table 1). A high proportion of the parents or guardians stated that the teeth of the children were cleaned daily (73–97% by age group), but plaque was still visible in a considerable portion of the children (23–50% by age group) (table 1). In all age groups, most of the children did not have visible caries, i.e. 94% among the 1- to 2-year-olds, 82% among the 2- to 3-year-olds and 77% among the 3- to 4-year-olds were caries free. Accordingly, the mean deft scores (numbers in parenthesis for means among children with caries) increased by age group from 0.16 (0.48) over 0.58 (1.75) to 0.93 (2.78) (table 1). The proportion of children with untreated cavities increased from 1.3 over 8.7 to 17.2% by increasing age group.

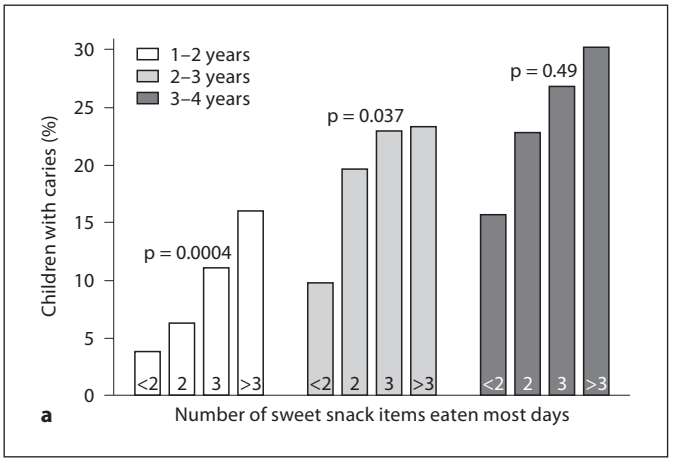
Visible Plaque and Income Are Highly Associated with Caries

A significantly higher proportion of the children with visible plaque had caries, and their age-standardized deft score was higher compared to plaque-free children (table 2). Similarly, children from homes with an income below the median, compared to those with an income at or above the median level, had more caries (table 2). The proportions of children with caries or mean deft scores did not differ between boys and girls, race or ethnicity, or parent/guardian education level groups (table 2).

Eating and Snacking Habits Are Associated with Caries

Nearly all children (97%) were reported to eat snacks most days, and 60% ate 1–2 sweet snack items most days (data not shown). The proportions of children with caries increased by increasing number of sweet items reported to be eaten most days (fig. 1a). There was less caries in children who drank milk compared with other drinks (non-sweetened or sweet) with the snacks (fig. 1b).

Caries was significantly more prevalent among children who ate chips most days, and their mean deft score was higher than in those who did not (table 3). Snacking on candies, cookies and ice cream was also associated with a higher proportion of children with caries. The mean deft score was significantly higher in children eating dry cereals and dried fruit. The most frequently consumed snack foods (fresh fruit, crackers and yoghurt), however, were not associated with caries (table 3).



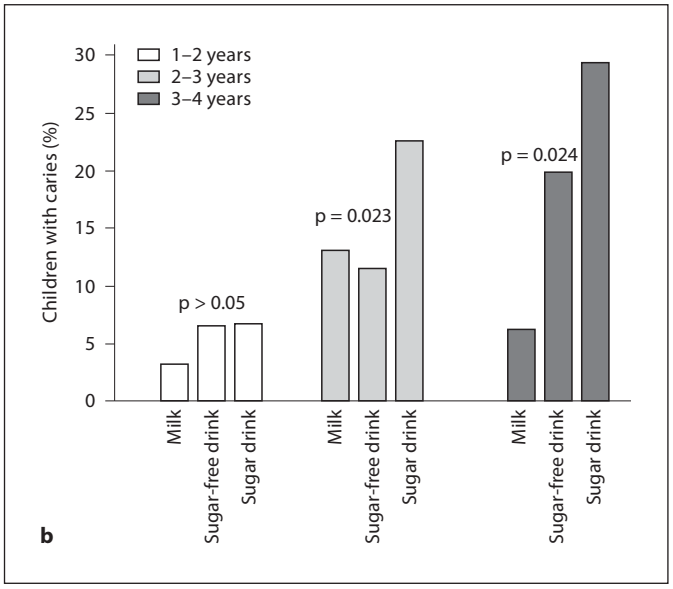


Fig. 1. p values from χ^2 analyses within each age category are indicated in the figures. p values for merged age categories were $p < 0.0001$ (**a**) and $p = 0.002$ (**b**). **a** Proportion of children with caries by number of sweet snack items eaten most days. **b** Proportion of children with caries by type of drink reported to be consumed with snack.

Children who were allowed to continue breast-feeding after falling asleep (mainly the same children reported to be breast-fed) had significantly higher deft values than those who were not (table 4). Allowing a bottle in bed was unrelated to caries prevalence in this population (table 4).

Multivariate Analysis

A multivariate PLS model with caries status was used to simultaneously evaluate the caries associations identified in the prior univariate analyses, and it included age

Table 3. Caries status by snack intake

Snack	Consumer	Numbers	Children with caries[1]		Caries experience[2]	
			%	p	deft	p
Chips	no	828	8.7	<0.0001	0.43 (0.31–0.56)	<0.0001
	yes	378	19.6		0.79 (0.64–0.94)	
Candies	no	1,008	10.3	<0.0001	0.53 (0.43–0.63)	0.217
	yes	198	21.2		0.68 (0.46–0.90)	
Ice cream	no	947	10.7	0.003	0.52 (0.42–0.62)	0.127
	yes	259	17.4		0.68 (0.50–0.86)	
Cookies	no	604	9.8	0.013	0.51 (0.37–0.65)	0.247
	yes	602	14.5		0.61 (0.49–0.73)	
Cereals + milk	no	862	10.7	0.014	0.51 (0.39–0.63)	0.124
	yes	344	15.7		0.66 (0.50–0.82)	
Cereals (dry)	no	678	11.2	0.279	0.46 (0.34–0.58)	0.008
	yes	528	13.3		0.70 (0.56–0.84)	
Dried fruit	no	1,042	11.7	0.286	0.52 (0.42–0.62)	0.039
	yes	164	14.6		0.78 (0.54–1.02)	
Popcorn	no	993	11.8	0.044	0.60 (0.48–0.78)	0.728
	yes	134	17.9		0.54 (0.29–0.59)	
Yoghurt	no	544	11.0	0.146	0.54 (0.40–0.68)	0.340
	yes	583	13.9		0.63 (0.49–0.77)	
Fresh fruit	no	264	11.4	0.676	0.54 (0.34–0.74)	0.837
	yes	942	12.3		0.56 (0.46–0.66)	
Crackers	no	376	10.4	0.214	0.47 (0.31–0.63)	0.941
	yes	830	12.9		0.60 (0.48–0.72)	
Others	no	841	12.4	0.698	0.56 (0.44–0.68)	0.941
	yes	363	11.6		0.55 (0.39–0.71)	

$p < 0.01$ considered statistically significant.
[1] Crude numbers tested by χ^2 test for differences in group proportions. $p < 0.01$ for statistical significance.
[2] Means (95% CI) standardized for age group by the GLM procedure (standardization for number of teeth had no further effect). Differences between groups tested by the Bonferroni post hoc test. $p < 0.05$ for statistical significance.

and number of teeth. In this model (ln-transformed deft) as y, and all variables describing intake of snacks, breast- and bottle-feeding, oral hygiene, tooth exposure time (child age, number of teeth), socioeconomic status (income, education) and gender were tested. The model had 2 significant components explaining 13% ($R^2 = 0.134$) and predicting 11% ($Q^2 = 0.109$) of the caries variation (fig. 2a, loading scatter plot; fig. 2b, correlation coefficient plot). The most influential variables in the model (all associated with high caries scores) were presence of plaque (VIP value = 2.39), descriptors for tooth exposure time (number of teeth: VIP value = 1.68; child age: VIP value = 1.50) and several dietary components: total intake of sugar-containing snacks (VIP value = 1.56), intake of chips (VIP value = 1.36), and chips intake with a sugar-containing drink (VIP value = 1.24). Snacking on candies had a borderline effect (VIP value = 0.96), whereas all other snacks had a minimal effect (VIP values <0.9). Models using a dichotomous caries variable provided the same results (data not shown).

PLS modelling also indicated that eating chips as a snack clustered with eating a high number of sweet snacks most days, eating candies, popcorn and ice cream (fig. 3, encircled factors all have VIP values >1.0).

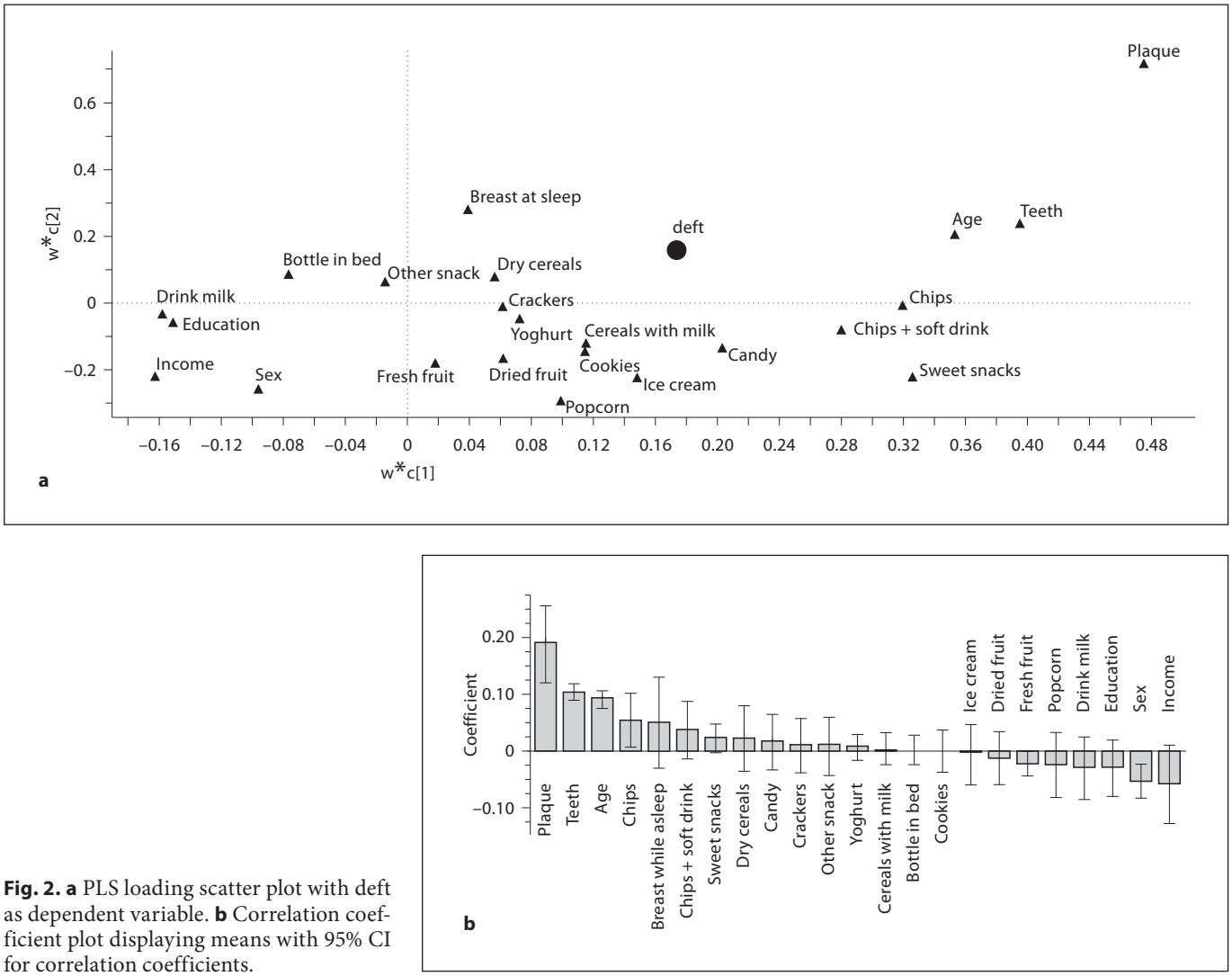


Fig. 2. a PLS loading scatter plot with deft as dependent variable. **b** Correlation coefficient plot displaying means with 95% CI for correlation coefficients.

Table 4. Caries status by eating habits

	Numbers	Children with caries[1]		Caries experience[2]	
		%	p	deft	p
Breast-feeding continued at sleep					
Yes (daily)	49	16.3	0.411	1.48 (1.01–1.95)	0.0003
Never/sparsely	906	13.1		0.61 (0.49–0.73)	
Bottle in bed at night or nap time					
Yes (daily)	429	9.6	0.047	0.53 (0.42–0.64)	0.233
Never/sparsely	773	13.5		0.64 (0.47–0.81)	

[1] Crude numbers tested by χ^2 test for differences in group proportions. $p < 0.01$ for statistical significance.
[2] Means (95% CI) standardized for age group by the GLM procedure (standardization for number of teeth had no further effect). Differences between groups tested by the Bonferroni post hoc test. $p < 0.05$ for statistical significance.

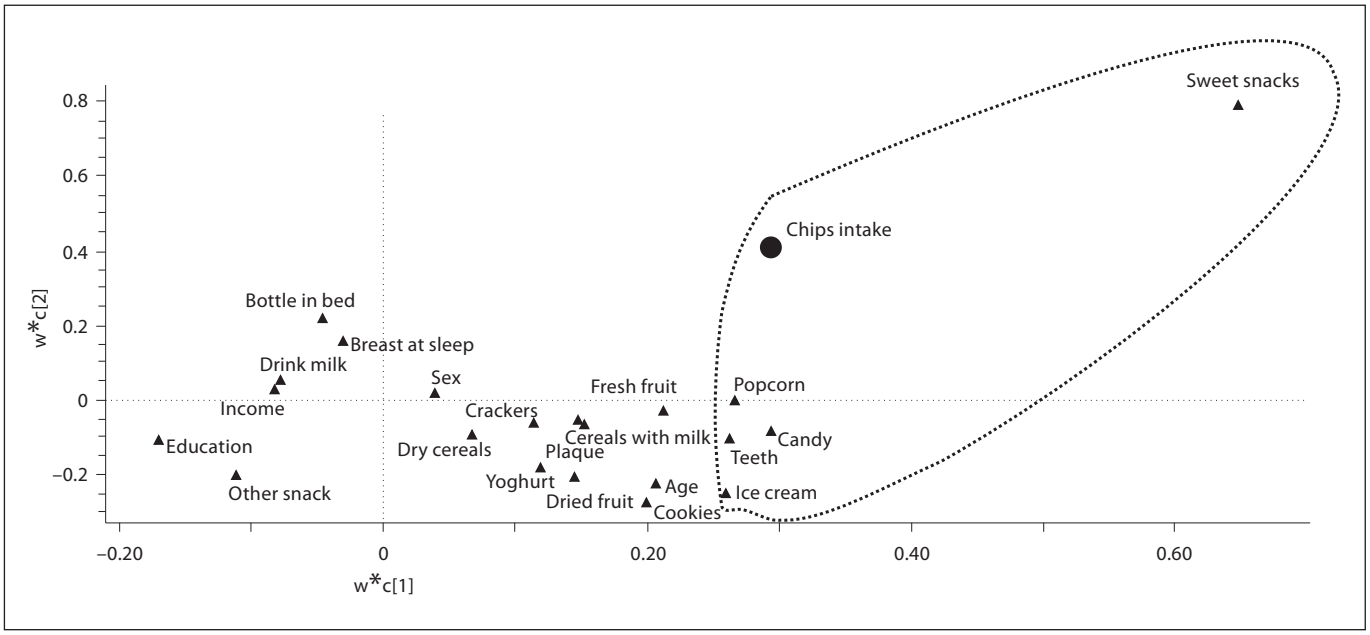


Fig. 3. PLS loading scatter plot modelled with intake of chips as the dependent variable.

Discussion

Repeated pH drop on the tooth surface from carbohydrate fermentation by dental biofilm bacteria is aetiologic to dental caries, but disease outcome is modulated by host and other factors. The relative role of various risk indicators, however, differs between populations [Mattos-Graner et al., 1998; Ramos-Gomez et al., 1999; Tada et al., 1999; Dini et al., 2000; Dasanyake and Caufield, 2002; Santos and Soviero, 2002; van Palenstein Helderman et al., 2006]. Notably, an association between sugar intake and caries development could not be demonstrated in many industrialized 'low-caries' countries [Garcia-Closas et al., 1997; Dye et al., 2004; Öhlund et al., 2007]. The present study shows that snacking on sucrose-containing products, as well as starch-containing chips and presence of plaque are associated with caries in children living in a low-socioeconomic-status, 'high-caries' area in an industrialized country (USA) [Kressin et al., 2009; Nunn et al., 2009].

The present study showed an association between consumption of chips and caries status not previously demonstrated in humans. This association might be explained by a direct caries-inducing effect of chips, or by chips intake clustering with other caries-promoting lifestyle factors. The present study design did not allow for such a distinction, and both aspects might well be involved. In vitro and in vivo studies have shown the pH-lowering effect of hydrolysed starch to be as rapid and deep as that of sucrose [Lingström et al., 1994], and in animal studies there has been shown a caries-inducing potential similar to that of sucrose [Mundorff-Shrestha et al., 1994]. In addition, animal studies have demonstrated that starch potentiates the cariogenic effect of sucrose [Ribeiro et al., 2005]. The PLS modelling using chips intake as the dependant variable, however, showed that children who were given chips as a snack most days also had a high intake of sweet snacks, ice cream and candies, illustrating a clustering of an unfavourable dietary patterns. Notably, chips intake was unrelated to presence of visible plaque.

Besides chips intake and indicators of possible total tooth exposure to cariogenic products – i.e. child age and number of teeth –, presence of visible plaque, number of sweet items used for daily snacking and low family income were independently associated with caries status in the multivariate PLS modelling in this study population. This conforms with earlier studies showing childhood caries was experienced more frequently in children who live under poor economic circumstances, belong to ethnic and racial minorities, have single mothers or have parents with low education, as well as with studies identifying plaque presence and sugar intake as risk indicators for caries [Harris et al., 2004; Vachirarojpisan et al., 2004; Gussy et al., 2006; Kanasi et al., 2010]. We found

PLS modelling to be a useful tool for identifying caries risk indicators as it allows a large number of explanatory variables to be examined simultaneously even though these variables covary [Jonasson et al., 2007; Kanasi et al., 2010].

The cariogenic potential of human milk with approximately 7% lactose has been questioned [Caplan et al., 2008; Mohebbi et al., 2008]. Several recent studies have shown that breast-feeding by itself does not increase the risk of caries in infants [Mohan et al., 1998; Iida et al., 2007; Mohebbi et al., 2008], with the possible exception of continued feeding after the child has fallen asleep [Roberts et al., 1993; Valaitis et al., 2000], as supported by the present data. Thus, children who were breast-fed after having fallen asleep tended to have more caries, but this finding would need confirmation by additional study since only a few children were breast-fed, and only the mean deft value, and not the proportion of breast-fed children with caries, was higher in breast-fed children. Bottle-feeding with sucrose-containing infant formulas, fruit soups or syrups, especially in bed, were reported to be risk indicators for caries in young children [Seow et al., 2009], but this was not confirmed in the present children. In contrast, consumption of cow's milk together with the suspected cariogenic snacks was associated with less caries than consumption of sugar-free or sugar-containing drinks.

The strengths of the present study were that (i) children were recruited from the general population in an area where caries in very young or preschool children is prevalent, and that recruitment was not restricted to those seeing a dentist, (ii) the children represented diverse races and ethnicities, and (iii) the number of children was high. Limitations were (i) the potential for recall bias in the survey information, (ii) the highly skewed low-caries distribution, and (iii) the lack of radiographs to examine teeth (which was not possible in the paediatrician's office). Bratthall [2000] has suggested the use of a Significant Caries Index, corresponding to data from the highest caries tertile to overcome a skewed distribution, but this was not applicable to this population since less than one third of the children had carious lesions.

Caries in early childhood is frequently observed, which preferentially should be addressed by preventive measures as treatment is costly and children with caries in early childhood are prone to high caries activity in the permanent dentition [Alm et al., 2007]. Caries risk is especially high in underprivileged or vulnerable groups [Nunn et al., 2009]. The present study suggests that good oral hygiene and promotion of healthy snacking remain targets for prevention in these children. In the present study, however, the risk indicators examined only explained 13.4% of the disease variation; thus, further studies using additional and new improved risk markers are still desirable.

Acknowledgements

The study was supported by the NIH/National Institute of Dental and Craniofacial Research grants U54 DE-014264, K24 DE-000419, DE-015847 (A.C.T.) and DE-019275 (M.E.N.), and by the County Council of Västerbotten, Sweden (I.J., P.L.H.).

References

Adair LS, Popkin BM: Are child eating patterns being transformed globally? Obes Res 2005; 13:1281–1299.

Alm A, Wendt LK, Koch G, Birkhed D: Prevalence of approximal caries in posterior teeth in 15-year-old Swedish teenagers in relation to their caries experience at 3 years of age. Caries Res 2007;41:392–398.

American Academy of Pediatric Dentistry: Definition of early childhood caries (ECC). Pediatr Dent 2008;27:13.

Azevedo TD, Bezerra AC, de Toledo OA: Feeding habits and severe early childhood caries in Brazilian preschool children. Pediatr Dent 2005;27:28–33.

Beltrán-Aguilar ED, Barker LK, Canto MT, Dye BA, Gooch BF, Griffin SO, Hyman J, Jaramillo F, Kingman A, Nowjack-Raymer R, Selwitz RH, Wu T, Centers for Disease Control and Prevention (CDC): Surveillance for dental caries, dental sealants, tooth retention, edentulism, and enamel fluorosis: United States, 1988–1994 and 1999–2002. MMWR Surveill Summ 2005;54:1–43.

Bratthall D: Introducing the Significant Caries Index together with a proposal for a new global oral health goal for 12-year-olds. Int Dent J 2000;50:378–384.

Briefel RR, Johnson CL: Secular trends in dietary intake in the Unites States. Annu Rev Nutr 2004;24:401–431.

Caplan LS, Erwin K, Lense E, Hicks J Jr: The potential role of breast-feeding and other factors in helping to reduce early childhood caries. J Public Health Dent 2008;68:238–241.

Dasanayake AP, Caufield PW: Prevalence of dental caries in Sri Lankan aboriginal Veddha children. Int Dent J 2002;52:438–444.

Dini EL, Holt RD, Beidi R: Caries and its association with infant feeding and oral health-related behaviours in 3–4-year-old Brazilian children. Community Dent Oral Epidemiol 2000;28:241–248.

Drury TF, Horowitz AM, Ismail AI, Maertens MP, Rozier RG, Selwitz RH: Diagnosing and reporting early childhood caries for research purposes: a report of a workshop sponsored by the National Institute of Dental and Craniofacial Research, the Health Resources and Services Administration, and the Health Care Financing Administration. J Public Health Dent 1999;59:192–197.

Dye BA, Shenkin JD, Ogden CL, Marshall TA, Levy SM, Kanellis MJ: The relationship between healthful eating practices and dental caries in children aged 2–5 years in the United States. J Am Dent Assoc 2004;135:55–66.

Garcia-Closas R, Garcia-Closas M, Serra-Majem L: A cross-sectional study of dental caries, intake of confectionary and foods rich in starch and sugars, and salivary counts of *Streptococcus mutans* in children in Spain. Am J Clin Nutr 1997;66:1257–1263.

Grindefjord M, Dahllöf G, Ekström G, Höjer B, Modéer T: Caries prevalence in 2.5-year-old children. Caries Res 1993;27:505–510.

Gussy MG, Waters EG, Walsh O, Kilpatrick NM: Early childhood caries: current evidence for aetiology. J Paediatr Child Health 2006;42:37–43.

Harris R, Nicoll AD, Adair PM, Pine CM: Risk factors for dental caries in young children: a systematic review of the literature. Community Dent Health 2004;21:71–85.

Iida H, Auinger P, Billings RJ, Weitzman M: Association between infant breastfeeding and early childhood caries in the United States. Pediatrics 2007;120:944–952.

Jonasson A, Eriksson C, Jenkinson H, Källestål C, Johansson I, Strömberg N: Innate immunity glycoprotein gp-340 variants may modulate human susceptibility to dental caries. BMC Infect Dis 2007;7:57.

Jose B, King NM: Early childhood caries lesions in preschool children in Kerala, India. Pediatr Dent 2003;25:594–600.

Kanasi E, Johansson I, Lu SC, Kressin NR, Nunn ME, Kent R Jr, Tanner ACR: Microbial risk markers for childhood caries in pediatricians' offices. J Dent Res 2010;89:378–383.

Kressin NR, Nunn ME, Singh H, Orner MB, Pbert L, Hayes C, Culler C, Glicken SR, Palfrey S, Geltman PL, Cadoret C, Henshaw MM: Pediatric clinicians can help reduce rates of early childhood caries: effects of a practice-based intervention. Med Care 2009;47:1121–1128.

Lingström P, Birkhed D, Ruben J, Arends J: Effect of frequent consumption of starchy food items on enamel and dentin demineralization and on plaque pH in situ. J Dent Res 1994;73:652–660.

Lingström P, van Houte J, Kashket S: Food starches and dental caries. Crit Rev Oral Biol Med 2000;11:366–380.

Marshall TA: The roles of meal, snack, and daily total food and beverage exposures on caries experience in young children. J Public Health Dent 2005;65:166–173.

Mattos-Graner RO, Zelante F, Line RC, Mayer MP: Association between caries prevalence and clinical, microbiological and dietary variables in 1.0- to 2.5-year-old Brazilian children. Caries Res 1998;32:319–323.

Milgrom P, Reisine S: Oral health in the United States: the postfluoride generation. Annu Rev Public Health 2000;21:403–436.

Milnes AR: Description and epidemiology of nursing caries. J Public Health Dent 1996;56:38–50.

Mohan A, Morse DE, O'Sullivan DM, Tinanoff N: The relationship between bottle usage/content, age, and number of teeth with mutans streptococci colonization in 6–24-month-old children. Community Dent Oral Epidemiol 1998;26:12–20.

Mohebbi SZ, Virtanen JI, Vahid-Golpayegani M, Vehkalahti MM: Feeding habits as determinants of early childhood caries in a population where prolonged breastfeeding is the norm. Community Dent Oral Epidemiol 2008;36:363–369.

Mundorff-Shrestha SA, Featherstone JD, Eisenberg AD, Cowles E, Curzon ME, Espeland MA, Shields CP: Cariogenic potential of foods. 2. Relationship of food composition, plaque microbial counts, and salivary parameters to caries in the rat model. Caries Res 1994;28:106–115.

Nunn ME, Dietrich T, Singh HK, Henshaw MM, Kressin NR: Prevalence of early childhood caries among very young urban Boston children compared with US children. J Public Health Dent 2009;69:156–162.

Öhlund I, Holgerson PL, Bäckman B, Lind T, Hernell O, Johansson I: Diet intake and caries prevalence in four-year-old children living in a low-prevalence country. Caries Res 2007;41:26–33.

Petersen PE, Esheng Z: Dental caries and oral health behaviour situation of children, mothers and schoolteachers in Wuhan, People's Republic of China. Int Dent J 1998;48:210–216.

Ramos-Gomez FJ, Tomar SL, Ellison J, Artiga N, Sintes J, Vicuna G: Assessment of early childhood caries and dietary habits in a population of migrant Hispanic children in Stockton, California. ASDC J Dent Child 1999;66:395–403.

Ribeiro CC, Tabchoury CP, del Bel Cury AA, Tenuta LM, Rosalen PL, Cury JA: Effect of starch on the cariogenic potential of sucrose. Br J Nutr 2005;94:44–50.

Roberts GJ, Cleatin-Jones PE, Fatti LP, Richardson BD, Sinwei RE, Margreaves JA, Williams S: Patterns of breast and bottle feeding and their association with dental caries in 1–4-year-old South African children. Community Dent Health 1993;10:405–413.

Santos AP, Soviero VM: Caries prevalence and risk factors among children aged 0 to 36 months. Pesqui Odontol Bras 2002;16:203–208.

Schroth RJ, Moore P, Brothwell DJ: Prevalence of early childhood caries in 4 Manitoba communities. J Can Dent Assoc 2005;71:567.

Seow WK, Clifford H, Battistutta D, Morawska A, Holcombe T: Case-control study of early childhood caries in Australia. Caries Res 2009;43:25–35.

Stecksen-Blicks C, Sunnegårdh K, Borssen E: Caries experience and background factors in 4-year-old children: time trends 1967–2002. Caries Res 2004;38:149–155.

Tada A, Ando Y, Hanada N: Caries risk factors among three-year-old children in Chiba, Japan. Asia Pac J Public Health 1999;11:109–112.

Vachirarojpisan T, Shinada K, Kawaguchi Y, Laungwechakan P, Somoke T, Detsomboonrat P: Early childhood caries in children aged 6–19 months. Community Dent Oral Epidemiol 2004;32:133–142.

Valaitis R, Hesch R, Passarelli C, Sheehan D, Sinton J: A systematic review of the relationship between breastfeeding and early childhood caries. Can J Public Health 2000;91:411–417.

van Palenstein Helderman WH, Soe W, van't Hof MA: Risk factors of early childhood caries in a Southeast Asian population. J Dent Res 2006;85:85–88.

Wennhall I, Matsson L, Schröder U, Twetman S: Caries prevalence in 3-year-old children living in a low socio-economic multicultural urban area in southern Sweden. Swed Dent J 2002;26:167–172.

RESEARCH REPORTS
Clinical

E. Kanasi[1,2,3], I. Johansson[4], S.C. Lu[1], N.R. Kressin[5,6,7], M.E. Nunn[5], R. Kent, Jr.[2,8], and A.C.R. Tanner[1,9]*

[1]Department of Molecular Genetics and [8]Department of Biostatistics, The Forsyth Institute, 140 The Fenway, Boston, MA 02115, USA; [2]Department of Oral Health Policy and Epidemiology and [9]Department of Oral Medicine, Infection and Immunity, Harvard School of Dental Medicine, Harvard University, Boston, MA 02115, USA; [3]Section of Oral Microbiology and [4]Section of Cariology, Department of Odontology, Faculty of Medicine, Umeå University, Umeå, 90187, Sweden; [5]Northeast Center for Research to Evaluate and Eliminate Dental Disparities (CREEDD), Goldman School of Dental Medicine, [6]Healthcare Disparities Research Program, Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, and [7]Boston VA Medical Center, Department of Veteran Affairs, Boston University, Boston, MA 02118, USA; *corresponding author, annetanner@forsyth.org

J Dent Res 89(4):378-383, 2010

Microbial Risk Markers for Childhood Caries in Pediatricians' Offices

ABSTRACT

Dental caries in pre-school children has significant public health and health disparity implications. To determine microbial risk markers for this infection, this study aimed to compare the microbiota of children with early childhood caries with that of caries-free children. Plaque samples from incisors, molars, and the tongue from 195 children attending pediatricians' offices were assayed by 74 DNA probes and by PCR to *Streptococcus mutans.* Caries-associated factors included visible plaque, child age, race, and snacking habits. Species were detected more frequently from tooth than tongue samples. *Lactobacillus gasseri* ($p < 0.01$), *Lactobacillus fermentum, Lactobacillus vaginalis*, and *S. mutans* with *Streptococcus sobrinus* (all $p < 0.05$) were positively associated with caries. By multifactorial analysis, the probiotic *Lactobacillus acidophilus* was negatively associated with caries. *Prevotella nigrescens* was the only species ($p < 0.05$) significantly associated with caries by the 'false discovery' rate. Analysis of the data suggests that selected *Lactobacillus* species, in addition to mutans streptococci, are risk markers for early childhood caries.

KEY WORDS: early childhood caries, *S. mutans, Lactobacillus.*

DOI: 10.1177/0022034509360010

Received April 23, 2008; Last revision October 6, 2009; Accepted October 21, 2009

A supplemental appendix to this article is published electronically only at http://jdr.sagepub.com/supplemental.

INTRODUCTION

Despite a high dental caries prevalence (28%) in US pre-school children (Beltran-Aguilar *et al.,* 2005), treatment is not universally available, particularly for disadvantaged children (Siegal *et al*., 2005). Thus, there is need for dental care through government and state programs (Edelstein, 2000; Lee *et al*., 2004), as well as for dental screenings, parental counseling, and fluoride application in pediatricians' offices (Lewis *et al*., 2000). Since dental caries is a bacterial infection, testing for caries microbial biomarkers could facilitate identifying children at greatest risk for caries and most in need of care.

The primary pathogens associated with dental caries are *Streptococcus mutans* and *Streptococcus sobrinus,* the mutans streptococci. Other caries-associated species include non-mutans *Streptococcus*, *Lactobacillus, Actinomyces, Bifidobacterium,* and *Veillonella* species (Van Houte, 1993). Studies of early childhood caries (ECC) microbiota using cultural (Marchant *et al.*, 2001) and molecular approaches (Corby *et al.*, 2005; Aas *et al.*, 2008) further expanded the range of species detected in caries. Several reports have also documented the detection of periodontal species in young children, including *Porphyromonas gingivalis, Aggregatibacter actinomycetemcomitans, Parvimonas micra* (*Peptostreptococcus micros*), *Tannerella forsythia,* and *Treponema denticola* (Kamma *et al.*, 2000; Tanner *et al.*, 2002b). Early colonization with periodontal pathogens probably reflects gingival inflammation, but could indicate risk of periodontitis in adulthood.

This study aimed primarily to determine microbial risk markers for early childhood caries that could be used in risk assessment for caries development. As a possible location for screening young children, microbial samples were obtained from pediatricians who served socio-economically disadvantaged children at increased risk for ECC. Positive associations between bacterial species with caries would suggest that they are caries biomarkers. These could be used in risk assessment to identify children most at risk for caries development and progression, and who would benefit from dental counseling and preventive protocols.

Table 1. Demographic and Clinical Characteristics of Study Population

	Caries-free n = 160	Caries n = 35	Caries-positive	p-value
Mean age (yrs) ± SD[a]	2.6 ± 0.9	2.8 ± 0.9		0.245[b]
Gender				
Male	74	22	23%	
Female	86	13	13%	0.075[c]
Race				
White	31	2	6%	
Black	122	28	19%	
Asian	7	5	42%	0.020[c]
Ethnicity				
Hispanic	21	2	9%	
Non-Hispanic	139	33	19%	0.383[d]
Child Dental Access				
Child has dentist	33	13	28%	
Child no dentist	127	22	15%	0.039[c]
Ever visited	22	13	37%	
Never visited	138	22	14%	< 0.001[c]
Snack Intake[e]				
Crackers	113	31	22%	
No crackers	47	4	8%	0.029[c]
Potato chips (crisps)	65	21	24%	
No potato chips	95	14	13%	0.037[c]
Cereal with milk	40	15	27%	
No cereal with milk	120	20	14%	0.033[c]
Plaque	73	29	28%	
Visible plaque	73	29	28%	
No plaque	87	6	6%	< 0.001[c]

[a] SD = Standard Deviation, [b] Student's *t* test, [c] Chi-square test, [d] Fisher's Exact test, [e] Most days of the week.

MATERIALS & METHODS

Study Population

Pre-school children were recruited in a cohort study from pediatric clinics (Boston Medical Center-Boston University, Floating Hospital-Tufts Medical Center, Boston, MA, USA). Inclusion criteria were that the child was from 1 to 6 yrs of age, had not taken antibiotics in the preceding 3 mos, and the parent or guardian was willing to consent to the child's clinical examination and microbial sampling. The study design, protocol, and informed consent were examined and approved by the Institutional Review Boards of the institutions involved. Children consenting to microbial samplings were from a larger cohort of children participating in a study to evaluate the efficacy of pediatrician training to reduce rates of early childhood caries (Kressin *et al.,* 2009).

Clinical Examination, Socio-demographic Information, and Bacterial Sampling

Children were examined by two trained and calibrated dental hygienists. The child's socio-demographic characteristics, diet, and oral health practices were collected by survey (Table 1). Presence of teeth and their status as sound, pre-cavitated, cavitated, filled, or sealed were recorded for each tooth (Drury *et al.*, 1999). Visible plaque was recorded on a 0-3 scale (no plaque, plaque covering a mean surface area of <1/3, 1/3 to <2/3, or >2/3 by tooth, respectively).

Plaque samples were taken with sterile toothpicks (Milgrom *et al*., 2000) from: (i) the gingival third of the labial coronal surfaces of incisors, (ii) the buccal and interproximal surfaces of molars, and (iii) across the midline of the tongue. In children with caries (pre-cavitated and cavitated), plaque was sampled from lesions. Samples were collected in 100 μL of 50 mM Tris-EDTA buffer (pH 7.6) and kept frozen at -70°C until microbial analysis.

Microbial Analyses

Plaque samples were analyzed by DNA probes to 74 species (Table 2) in a checkerboard assay as previously described (Tanner *et al.*, 2002a; Socransky *et al.*, 2004). Samples were denaturated, neutralized, and fixed onto a nylon membrane, the last lanes of which were used for quantitation, containing all probe species mixed as a DNA standard at 10^5 and 10^6 bacterial cell equivalents.

For *S. mutans* PCR detection, samples were treated with proteinase K (Dewhirst *et al*., 2000). The assay was run with *S. mutans*-specific reverse primer 5′-ACT CCA GAC TTT CCT GAC CG-3′ and forward universal primer 5′-GAG TTT GAT YMT GGC TCA G-3′ (Paster *et al.,* 1998). The amplification products were visualized with 1% agarose gels. If an amplicon was not detected, a universal PCR (primers: forward as above, reverse 5′-AAG GAG GTG WTC CAR CC-3′) was run to verify the presence of sample DNA.

Data Analyses

Decayed (d), comprising pre-cavitated and cavitated lesions, and filled (f), not including sealants, primary teeth (t) for each child were summarized by a decayed-filled tooth score (dft). Plaque was dichotomized to 'presence' and 'absence', since half (48%) of the observations were 'no plaque' and half (52%) were 'plaque detected'. Species detection was at $\geq 10^5$ cells with the use of DNA probes (equivalent to the lower DNA internal standard), or a positive PCR reaction. Tooth and tongue samples were matched by child. We used Student's *t* test to compare differences in means,

Table 2. Microbiota of Incisor, Molar, and Tongue Samples of Pre-school Children

Bacterial Species[a]	Percent Detection Frequency 10^5		
	Incisor	Molar	Tongue
Streptococcus mutans	32	28	8
Streptococcus sobrinus	40	39	18
S. mutans with *S. sobrinus*	20	19	4
Streptococcus cristatus	53	56	14
Streptococcus intermedius	43	41	17
Streptococcus constellatus	42	40	9
Lactobacillus vaginalis	42	41	14
Lactobacillus fermentum	34	27	13
Lactobacillus gasseri	34	32	13
Lactobacillus rhamnosus	23	25	9
Lactobacillus acidophilus	46	41	13
Lactobacillus plantarum	21	23	7
Lactobacillus reuteri	16	19	8
Bifidobacterium dentium	54	51	16
Rothia dentocariosa	65	54	13
Actinomyces odontolyticus	87	81	50
Actinomyces meyeri	49	44	17
Actinomyces israelii	68	64	15
Eubacterium saburreum	52	50	6
Eubacterium saphenum[b]	12	13	7
Eubacterium limosum	31	30	11
Eubacterium timidum	36	32	19
Eubacterium brachy	20	19	9
Eubacterium nodatum	28	20	9
Parvimonas micra	50	51	27
Filifactor alocis	64	53	30
Prevotella nigrescens	39	44	9
Porphyromonas endodontalis	23	21	4
Porphyromonas gingivalis	12	13	3
Tannerella forsythia	27	22	10
Capnocytophaga gingivalis	40	45	7
A. actinomycetemcomitans	18	19	3
Kingella oralis	85	81	38
Campylobacter gracilis	23	33	5
F. nucleatum subsp. *polymorphum*	53	59	18
Leptotrichia buccalis	62	65	15
Veillonella parvula	65	71	28
Selenomonas sputigena	42	43	19
Selenomonas noxia	54	49	7
Selenomonas flueggei	40	38	12
Treponema denticola[b]	17	20	10

[a] Species in Table 2 differed significantly between caries-affected and caries-free children (Fig. 1) or were significant in the PLS analysis (Fig. 2).

[b] Species with similar detection between incisors and tongue (McNemar's test, $p > 0.05$).

All tested species are in the Appendix Table. All species were detected at similar frequencies from incisors and molars and more frequently from teeth than from the tongue (McNemar's test, $p \leq 0.05$).

chi-square test for proportions between socio-demographics, plaque, species detection, and caries, and McNemar's test for species detection at paired intra-oral sites. A p-value ≤ 0.05 was considered significant. Results were also adjusted for multiple comparisons by the false-discovery rate ($\alpha = 0.05$) (Benjamini and Hochberg, 1995). We used regression analysis to evaluate child caries extent with age and plaque and their interactions, and model sensitivity analysis was performed by overdispersed Poisson regression. Statistical analyses were performed with SPSS® software.

Multivariate analysis was performed with Partial Least Squares (PLS) modeling (SIMCA P, Umetrics, Umeå, Sweden) (Hellberg *et al.,* 1986). PLS is a linear model that detects correlations between matrices of each independent (x) variable and outcome (y) variables to generate model R^2- (estimated model explanation) and Q^2- (estimated model prediction) values, and variable Importance in Projection (VIP) values, which, if >1.0, indicates influential and, if VIP $\geq$ 1.5, highly influential x-variables. Separate and combined microbial and survey data were modeled on caries (dft) by logarithmic transformation [$\log_{10}$(dft) after input of 0.01 for 0] and dichotomization (0 for dft = 0 and 1 for dft > 0) of dft values. SIMCA P auto-centering and scaling to unit variance were applied to all variables, and leave-one-out cross-validation was used. Generally, R^2- and Q^2- values should not differ by more than 0.2.

RESULTS

Demographic and Clinical Characteristics

The population sample was comprised of 195, predominantly Black (77%), children (Table 1). For most children, parental education was $\leq$ 12 yrs (57%), and annual household income was < $20,000 (46%). Caries was detected in 18% of the children, mainly affecting Asian (42%) and Black (19%) children ($p < 0.05$) (Table 1). Survey variables associated with caries were: previous child dental visits, crackers, potato chips, and cereal consumption (all $p < 0.05$), and visible plaque ($p < 0.001$) (Table 1). In the children without visible plaque, caries (6%) was unrelated to age. In the children with plaque (52% of children), caries (28%) was related to age ($p = 0.02$). Caries extent (dft) increased with plaque presence and higher age ($p < 0.05$, data not shown). Child age ($p = 0.0041$) and detectable plaque ($p < 0.0001$) were also significant after sensitivity analysis by overdispersed Poisson regression.

Microbiology

DNA probe data from matched tooth and tongue samples (Table 2 and Appendix Table) showed a consistently higher species detection frequency from teeth compared with the tongue. Species were detected at similar frequencies from incisor and molar samples ($p = 0.09$-1.00). The most frequently detected species (>60% of children) included *Streptococcus* and *Actinomyces* species, *Rothia dentocariosa, Filifactor alocis,* and *Veillonella parvula*. Caries-associated species ($p \leq 0.05$) from molar plaque samples (Fig. 1 and Appendix Fig) included: *S. mutans*, *S. mutans* with *S. sobrinus* ($p = 0.03$), *Streptococcus intermedius, Lactobacillus vaginalis, Lactobacillus fermentum, Lactobacillus gasseri* ($p < 0.01$), *Actinomyces odontolyticus*, *Actinomyces israelii,* and *V. parvula* ($p < 0.01$). Subgingival species detected more frequently from caries-affected children ($p \leq 0.05$) included *Eubacterium brachy, P. micra, F. alocis, P. gingivalis* ($p < 0.01$), *Prevotella nigrescens*, *Porphyromonas endodontalis*, *A. actinomycetemcomitans,* and *Leptotrichia buccalis* ($p < 0.05$). *P. nigrescens* remained significant ($p < 0.05$) after multiple-comparisons adjustment.

S. mutans was detected by PCR in 41% molar, 32% incisor, and 14% tongue plaque samples from 169 children with valid samples from all sites (data not shown). Detection frequencies of *S. mutans* from caries-affected and caries-free children,

respectively, were: molar, 69% and 35%; incisor, 63% and 25% (both $p < 0.001$); and tongue, 22% and 12% ($p = 0.167$). *S. mutans* was associated with children with plaque compared with plaque-free children in molar (51% and 31%) ($p < 0.01$), incisor (42% and 21%) ($p < 0.01$), and tongue (20% and 6%) ($p < 0.05$) samples, respectively.

Initial multivariate analyses (PLS) identified age and number of teeth at risk as influencing caries detection. Modeling, performed in age groups (< 2 yrs, ≥ 2 to < 3 yrs, and ≥ 3 yrs of age) with dft as the dependent variable, indicated that bacterial detection and the presence of visible plaque yielded statistically significant models for the three age groups ($R^2 = 0.78$, $Q^2 = 0.21$; $R^2 = 0.54$, $Q^2 = 0.15$; and $R^2 = 0.53$, $Q^2 = 0.14$, respectively). There were no significant models for caries with the survey data alone, and combining microbial and survey data improved the model values only marginally.

When data from all children were analyzed, plaque and *S. mutans* detection by PCR were highly influential variables for higher dft scores (VIP > 1.5) (data not shown). *L. vaginalis, A. odontolyticus*, *A. israelii, P. nigrescens*, *Prevotella loescheii*, *P. gingivalis, A. actinomycetemcomitans,* and *V. parvula* were consistently influential in children with higher dft (VIP > 1.0). In contrast, *Lactobacillus acidophilus*, *Lactobacillus reuteri*, *Eubacterium saburreum, Eubacterium nodatum,* and *Campylobacter gracilis* were consistently influential in children with lower dft (VIP > 1.0).

When data were restricted to children with visible plaque, *S. mutans* by PCR was highly associated with caries (model $R^2 = 0.48$, $Q^2 = 0.23$) (VIP>1.5, Fig. 2). *A. israelii, Eubacterium timidum, A. actinomycetemcomitans,* and *V. parvula* were influential with caries extent (higher dft) (VIP > 1.0-1.3), whereas *L. gasseri, B. dentium, A. odontolyticus,* and *P. nigrescens* were borderline influential (VIP = 0.9). In children with plaque, the significant negative influence of *L. acidophilus*, *E. saburreum,* and *E. nodatum* on caries was confirmed (VIP > 1.5), and several additional species exhibited highly influential negative associations with caries (VIP = 1.0-1.4, Fig. 2). Caries-associated species explained 50% of caries extent.

Log-transformed and dichotomized dft gave similar results in all PLS models.

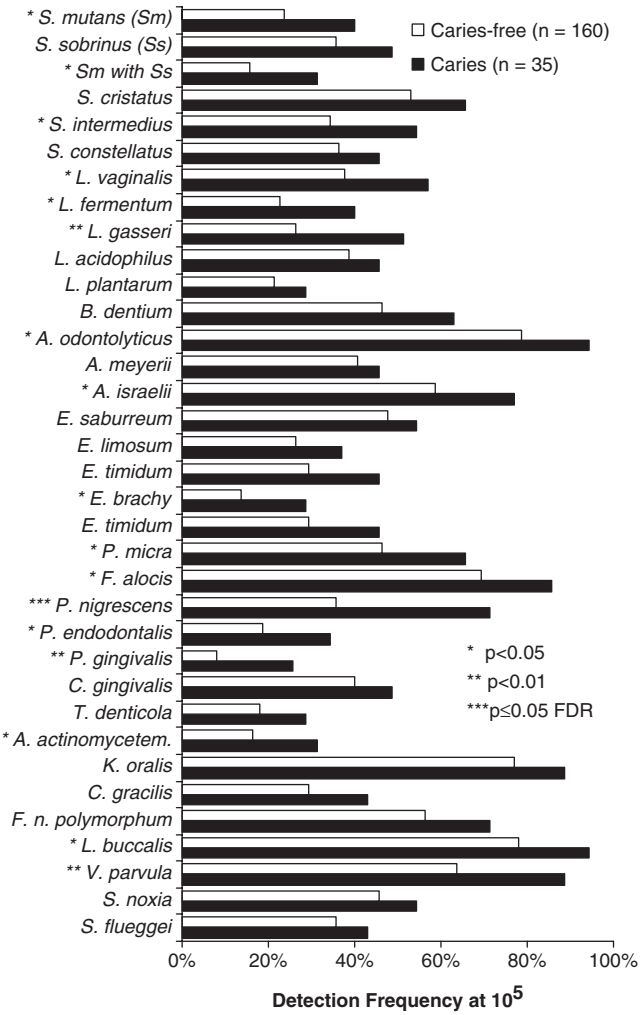


Figure 1. Microbiota of caries-affected and caries-free children. Bacterial species with > 5% difference in detection between caries-affected and caries-free children are in the same order as in Table 2. Most bacterial species assayed were detected more frequently (*$p \leq 0.05$ or **$p \leq 0.01$) from children with caries than from caries-free children. *P. nigrescens* association remained significant after adjustment for multiple comparisons by the false-discovery rate (***$p \leq 0.05$ FDR). *A. actinomycet.* is *Aggregatibacter actinomycetemcomitans,* and *F.n. polymorphum* is *Fusobacterium nucleatum* subspecies *polymorphum*.

DISCUSSION

Microbiology sampling was successfully integrated into pediatric practices serving low-income populations as a means of seeking microbial biomarkers for ECC. This study expanded microbial risk markers for ECC beyond mutans streptococci to include selected *Lactobacillus* species. Non-microbial caries risk factors included visible plaque, child age, race, and snacking habits. Detection of subgingival/periodontal species in children was confirmed. While the PLS statistical models showed low predictive power, R^2-$Q^2 > 0.2$, and suggested cautious interpretation of findings, overall the data are generally consistent with previous reports, as discussed below.

S. sobrinus detection with *S. mutans* showed high associations with ECC, which is consistent with studies correlating this species combination with caries development (Seki *et al.*, 2006) and higher caries prevalence than *S. mutans* alone (Okada *et al.*, 2005). Among the non-mutans streptococci, *S. intermedius* was associated with ECC, as previously reported in a different population of young children (Tanner *et al.*, 2002b).

Caries-associated *Lactobacillus* species in the current report included *L. gasseri*, previously detected in childhood and adult caries (Munson *et al.*, 2004; Corby *et al.*, 2005), *L. fermentum,* detected in childhood caries (Marchant *et al.*, 2001; Aas *et al.*, 2008), and *L. vaginalis,* reported in caries-active young women's saliva (Caufield *et al.,* 2007). Most of these *Lactobacillus* species were also detected by molecular methods from adult carious dentin (Chhour *et al.*, 2005). In contrast, the *Lactobacillus* species *L. rhamnosus, L. acidophilus*, *L. plantarum*, and *L. reuteri*, which are used in probiotic infant enteric and/or oral anti-caries therapy (Savino *et al.,* 2007; Twetman and Stecksén-Blicks, 2008), showed negative associations with caries. *A. israelii* and *A. odontolyticus* were detected more frequently in caries, as previously reported (Marchant *et al.,* 2001; Tanner *et al.*, 2002b).

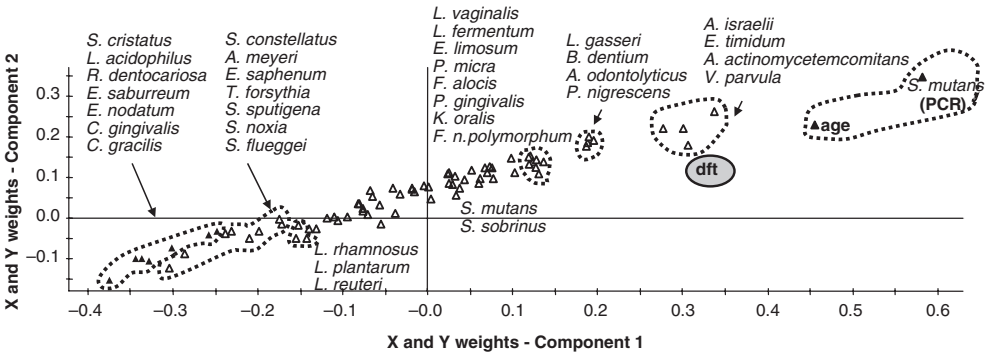


Figure 2. Scatter plot of Partial Least Squares (PLS) weights w*c for bacterial data modeled on caries prevalence (dft) in children with visible plaque (n = 102). For a given PLS model, one vector of X-weights w^*_a and one vector of Y-weights c_a are obtained for each model component (a). The X- and Y-weights are plotted together for the first two components (w*c [1] and w*c [2]). Variable importance in projection (VIP) indicates how influential the outcome (microbial) variables are in explaining the variable of interest (decayed filled teeth-dft). VIP > 1.0 is influential, and VIP ≥ 1.5 is highly influential. In this plot, species located close to the outcome variable dft were positively associated with dft, whereas those in the opposite direction represent species negatively associated with dft. Encircled ▲ indicate positions of highly influential species (VIP > 1.5), and encircled △ positions of influential (VIP 1.0-1.4) or borderline influential species (VIP 0.7-0.9). Non-labeled, not encircled △ indicate positions of non-influential species. *S. mutans* and *S. sobrinus* (when detected by checkerboard) were positively associated with caries, but were non-influential in this model.

Subgingival periodontal species were also detected in these young children, particularly from the molar plaque samples that included gingival bacteria. Subgingival species detected included *A. actinomycetemcomitans, P. micra, P. gingivalis, T. forsythia,* and *T. denticola,* as has previously been reported in young children (Könönen *et al.*, 1992; Tanner *et al.*, 2002b), at detection frequencies consistent with cultural assay (Kamma *et al.*, 2000). Other subgingival species associated with adult periodontitis (Kumar *et al.*, 2005), but infrequently reported in pre-school children, included *E. brachy*, *Eubacterium saphenum*, *F. alocis*, and *Selenomonas flueggei.* Colonization of these species probably reflected gingivitis (not consistently measured in this study), but may pose risk for future periodontitis.

Increased species detection in caries may be a reflection of larger plaque samples. Nonetheless, comparable caries-associated species were detected in all children compared with children with visible plaque. In other studies, visible plaque was strongly associated with caries (Wennhall *et al.*, 2002; Mohebbi *et al.*, 2006), and was an indicator for future caries development in young children (Alaluusua and Malmivirta, 1994).

For microbial screening, plaque sampling should be easy and rapid. Similar species detection rates between incisors and molars suggest that either tooth site could be sampled, in contrast to tongue samples, which exhibited lower detection frequencies of DNA probe species. Lower species detection from the tongue than from the teeth suggests that screening for caries pathogens in tongue samples is less sensitive than screening samples from teeth, and may result in false-negative data.

Caries prevalence in this study was similar to US national levels (Beltran-Aguilar *et al.*, 2005). Increased caries experience in Asian and Black compared with White children has also been reported, with Asians having the highest caries prevalence (30%) and untreated caries (49%) (Shiboski *et al*., 2003). There was more caries (dft) in the older children, reflecting the longer time teeth were exposed to caries-risk factors. Other caries-associated risk factors were consistent with previous findings. Frequent snacking was associated with caries. High-starch foods like potato chips and crackers stay in the mouth longer and aid in increased acid production, contributing to caries development (Kashket *et al.*, 1996). The association of ECC and having a dentist might seem counterintuitive. Families with children with cavities and possibly pain, however, are more likely to seek treatment than are dentally healthy children.

In conclusion, children in disadvantaged populations were sampled to evaluate microbial risk biomarkers for caries, indicating the feasibility of accessing these children while attending routine and emergency visits in pediatricians' offices. Using univariate and multivariate approaches, we associated selected *Lactobacillus* species, in addition to mutans streptococci, with early childhood caries. Future study will indicate if these species are risk indicators for future caries, and if they can be used to select children most in need of preventive and treatment regimens.

ACKNOWLEDGMENTS

We thank Harpreet Singh and Cindy Cadoret for subject recruitment, clinical measurements, and sampling. This investigation was supported by USPHS grants U54 DE-014264, DE-015847, DE-007151-18, DE-007327-07, and K24 DE000419 from the National Institute of Dental and Craniofacial Research, National Institutes of Health, Bethesda, MD 20892, USA, and by a Medical Faculty PhD grant, Umeå University, Sweden.

REFERENCES

Aas JA, Griffen AL, Dardis SR, Lee AM, Olsen I, Dewhirst FE, *et al*. (2008). Bacteria of dental caries in primary and permanent teeth in children and young adults. *J Clin Microbiol* 46:1407-1417.

Alaluusua S, Malmivirta R (1994). Early plaque accumulation—a sign for caries risk in young children. *Community Dent Oral Epidemiol* 22(5 Pt 1):273-276.

Beltran-Aguilar ED, Barker LK, Canto MT, Dye BA, Gooch BF, Griffin SO, *et al*. (2005). Surveillance for dental caries, dental sealants, tooth retention, edentulism, and enamel fluorosis—United States, 1988-1994 and 1999-2002. *MMWR Surveill Summ* 54:1-43.

Benjamini Y, Hochberg Y (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Statist Soc B* 57:289-300.

Caufield PW, Li Y, Dasanayake A, Saxena D (2007). Diversity of lactobacilli in the oral cavities of young women with dental caries. *Caries Res* 41:2-8.

Chhour KL, Nadkarni MA, Byun R, Martin FE, Jacques NA, Hunter N (2005). Molecular analysis of microbial diversity in advanced caries. *J Clin Microbiol* 43:843-849.

Corby PM, Lyons-Weiler J, Bretz WA, Hart TC, Aas JA, Boumenna T, *et al*. (2005). Microbial risk indicators of early childhood caries. *J Clin Microbiol* 43:5753-5759.

Dewhirst FE, Tamer MA, Ericson RE, Lau CN, Levanos VA, Boches SK, *et al*. (2000). The diversity of periodontal spirochetes by 16S rRNA analysis. *Oral Microbiol Immunol* 15:196-202.

Drury TF, Horowitz AM, Ismail AI, Maertens MP, Rozier RG, Selwitz RH (1999). Diagnosing and reporting early childhood caries for research purposes. A report of a workshop sponsored by the National Institute of Dental and Craniofacial Research, the Health Resources and Services Administration, and the Health Care Financing Administration. *J Public Health Dent* 59:192-197.

Edelstein BL (2000). Access to dental care for Head Start enrollees. *J Public Health Dent* 60:221-229.

Hellberg S, Sjöström M, Wold S (1986). The prediction of bradykinin potentiating potency of pentapeptides. An example of a peptide quantitative structure-activity relationship. *Acta Chem Scand B* 40:135-140.

Kamma JJ, Diamanti-Kipioti A, Nakou M, Mitsis FJ (2000). Profile of subgingival microbiota in children with primary dentition. *J Periodontal Res* 35:33-41.

Kashket S, Zhang J, Van Houte J (1996). Accumulation of fermentable sugars and metabolic acids in food particles that become entrapped on the dentition. *J Dent Res* 75:1885-1891.

Könönen E, Asikainen S, Jousimies-Somer H (1992). The early colonization of Gram-negative anaerobic bacteria in edentulous infants. *Oral Microbiol Immunol* 7:28-31.

Kressin RN, Nunn ME, Singh H, Orner MB, Pbert L, Hayes C, *et al*. (2009). Effects of a practice based intervention on pediatric providers' counseling about children's risk for dental decay. *Medical Care* 11:1121-1128.

Kumar PS, Griffen AL, Moeschberger ML, Leys EJ (2005). Identification of candidate periodontal pathogens and beneficial species by quantitative 16S clonal analysis. *J Clin Microbiol* 43:3944-3955.

Lee JY, Rozier RG, Norton EC, Kotch JB, Vann WF Jr (2004). The effects of the Women, Infants, and Children's Supplemental Food Program on dentally related Medicaid expenditures. *J Public Health Dent* 64:76-81.

Lewis CW, Grossman DC, Domoto PK, Deyo RA (2000). The role of the pediatrician in the oral health of children: a national survey. *Pediatrics* 106:E84.

Marchant S, Brailsford SR, Twomey AC, Roberts GJ, Beighton D (2001). The predominant microflora of nursing caries lesions. *Caries Res* 35:397-406.

Milgrom P, Riedy CA, Weinstein P, Tanner AC, Manibusan L, Bruss J (2000). Dental caries and its relationship to bacterial infection, hypoplasia, diet, and oral hygiene in 6- to 36-month-old children. *Community Dent Oral Epidemiol* 28:295-306.

Mohebbi SZ, Virtanen JI, Vahid-Golpayegani M, Vehkalahti MM (2006). Early childhood caries and dental plaque among 1-3-year-olds in Tehran, Iran. *J Indian Soc Pedod Prev Dent* 24:177-181.

Munson MA, Banerjee A, Watson TF, Wade WG (2004). Molecular analysis of the microflora associated with dental caries. *J Clin Microbiol* 42:3023-3029.

Okada M, Soda Y, Hayashi F, Doi T, Suzuki J, Miura K, *et al*. (2005). Longitudinal study of dental caries incidence associated with *Streptococcus mutans* and *Streptococcus sobrinus* in pre-school children. *J Med Microbiol* 54(Pt 7):661-665.

Paster BJ, Bartoszyk IM, Dewhirst FE (1998). Identification of oral streptococci using PCR-based, reverse-capture, checkerboard hybridization. *Methods Cell Sci* 20:223-231.

Savino F, Pelle E, Palumeri E, Oggero R, Miniero R (2007). *Lactobacillus reuteri* (American Type Culture Collection Strain 55730) versus simethicone in the treatment of infantile colic: a prospective randomized study. *Pediatrics* 119:e124-e130.

Seki M, Yamashita Y, Shibata Y, Torigoe H, Tsuda H, Maeno M (2006). Effect of mixed mutans streptococci colonization on caries development. *Oral Microbiol Immunol* 21:47-52.

Shiboski CH, Gansky SA, Ramos-Gomez F, Ngo L, Isman R, Pollick HF (2003). The association of early childhood caries and race/ethnicity among California preschool children. *J Public Health Dent* 63:38-46; *erratum* in *J Public Health Dent* 63:264, 2003.

Siegal MD, Marx ML, Cole SL (2005). Parent or caregiver, staff, and dentist perspectives on access to dental care issues for Head Start children in Ohio. *Am J Public Health* 95:1352-1359.

Socransky SS, Haffajee AD, Smith C, Martin L, Haffajee JA, Uzel NG, *et al*. (2004). Use of checkerboard DNA-DNA hybridization to study complex microbial ecosystems. *Oral Microbiol Immunol* 9:352-362.

Tanner AC, Milgrom PM, Kent R Jr, Mokeem SA, Page RC, Liao SI, *et al*. (2002a). Similarity of the oral microbiota of pre-school children with that of their caregivers in a population-based study. *Oral Microbiol Immunol* 17:379-387; *erratum* in *Oral Microbiol Immunol* 18:338, 2003.

Tanner AC, Milgrom PM, Kent R Jr, Mokeem SA, Page RC, Riedy CA, *et al*. (2002b). The microbiota of young children from tooth and tongue samples. *J Dent Res* 81:53-57.

Twetman S, Stecksén-Blicks C (2008). Probiotics and oral health effects in children. *Int J Paediatr Dent* 18:3-10.

Van Houte J (1993). Microbiological predictors of caries risk. *Adv Dent Res* 7:87-96.

Wennhall I, Matsson L, Schroder U, Twetman S (2002). Caries prevalence in 3-year-old children living in a low socio-economic multicultural urban area in southern Sweden. *Swed Dent J* 26:167-172.

REVIEW ARTICLE

Pharmacoeconomics 2010; 28 (7): 521-538
1170-7690/10/0007-0521/$49.95/0

Cost Effectiveness of Venous Thromboembolism Pharmacological Prophylaxis in Total Hip and Knee Replacement

A Systematic Review

Alok Kapoor,[1] *Warren Chuang*,[2] *Nila Radhakrishnan*,[1] *Kenneth J. Smith*,[3] *Dan Berlowitz*,[4,5] *Jodi B. Segal*,[6] *Jeffrey N. Katz*[7] and *Elena Losina*[8,9]

1 Hospital Medicine Unit, Boston University School of Medicine, Boston, Massachusetts, USA
2 Hospitalist Medicine Group, Massachusetts General Hospital, Boston, Massachusetts, USA
3 Section of Decision Sciences and Clinical Systems Modelling, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, USA
4 Center for Health Quality, Outcomes, and Economic Research, Edith Nourse Rogers Memorial VA Hospital, Bedford, Massachusetts, USA
5 Boston University School of Medicine, Boston, Massachusetts, USA
6 Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA
7 Department of Orthopaedic Surgery and Division of Rheumatology, Immunology, and Allergy, and Orthopaedic and Arthritis Center for Outcomes Research, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA
8 Department of Orthopaedic Surgery, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA
9 Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, USA

Contents

Abstract

Total hip and knee replacements (THR and TKR) are high-risk settings for venous thromboembolism (VTE). This review summarizes the cost effectiveness of VTE prophylaxis regimens for THR and TKR. We searched MEDLINE (January 1997 to October 2009), EMBASE (January 1997 to June 2009) and the UK NHS Economic Evaluation Database (1997 to October 2009).

We analysed recent cost-effectiveness studies examining five categories of comparisons: (i) anticoagulants (warfarin, low-molecular-weight heparin [LMWH] or fondaparinux) versus acetylsalicylic acid (aspirin); (ii) LMWH versus warfarin; (iii) fondaparinux versus LMWH; (iv) comparisons with new oral anticoagulants; and (v) extended-duration (≥3 weeks) versus short-duration (<3 weeks) prophylaxis. We abstracted information on cost and effectiveness for each prophylaxis regimen in order to calculate an incremental cost-effectiveness ratio. Because of variations in effectiveness units reported and horizon length analysed, we calculated two cost-effectiveness ratios, one for the number of symptomatic VTE events avoided at 90 days and the other for QALYs at the 1-year mark or beyond.

Our search identified 33 studies with 67 comparisons. After standardization, comparisons between LMWH and warfarin were inconclusive, whereas fondaparinux dominated LMWH in nearly every comparison. The latter results were derived from radiographic VTE rates. Extended-duration prophylaxis after THR was generally cost effective. Small numbers prohibit conclusions about aspirin, new oral anticoagulants or extended-duration prophylaxis after TKR.

Fondaparinux after both THR and TKR and extended-duration LMWH after THR appear to be cost-effective prophylaxis regimens. Small numbers for other comparisons and absence of trials reporting symptomatic endpoints prohibit comprehensive conclusions.

In 2005, there were 580 000 total hip or knee replacements (THR or TKR) performed in the US,[1] and that number is projected to increase to 4.5 million by 2030.[2] Although THR and TKR are generally safe procedures,[3] they have been identified[4] as high-risk events for venous thromboembolism (VTE), which includes deep venous thrombosis (DVT) and pulmonary embolism (PE). For almost 20 years, physicians have been offering pharmacological prophylaxis to patients undergoing THR and TKR; however, uncertainty exists about the optimal pharmacological regimen for prophylaxis.

Guidelines[4] published by the American College of Chest Physicians (ACCP) in 2008 support using potent anticoagulant regimens with agents such as fondaparinux sodium, low-molecular-weight heparin (LMWH) and warfarin (target international normalized ratio [INR] 2–3), and discourage acetylsalicylic acid (aspirin) therapy alone. In contrast, guidelines[5] by the American Academy of Orthopaedic Surgeons (AAOS) support the use of aspirin or a lower-potency warfarin regimen (INR < 2) in addition to LMWH and fondaparinux, stating that the latter agents do not offer increased protection against PE but substantially raise the rate of bleeding complications. New oral anticoagulants such as dabigatran etexilate, a direct thrombin inhibitor, are expected to gain US FDA approval within the next several months and it is anticipated that they too will be supported by the above professional societies.

Several studies have attempted to address these risk-benefit and also cost issues using decision analysis methodology. Some studies[6-8] indicate that the newer regimens are cost effective in preventing VTE, while others[9] do not. Individual study results vary depending on the setting, economic perspective (e.g. groups for which cost and effects will be aggregated – patients, payers or others), horizon (time course over which cost and effectiveness information was assessed) and effectiveness outcome analysed (e.g. VTE events averted, life-years gained, QALYs gained). Measuring effectiveness in QALYs, particularly over a horizon of ≥1 year, permits comparison of cost effectiveness of interventions across diseases, but some authors may choose not to measure QALYs because their focus resides in the economics related to the period immediately following surgery. To more meaningfully compare VTE prophylaxis regimens, we systematically reviewed recently published studies that evaluated the cost effectiveness of the different pharmacological options in patients undergoing THR and TKR. We abstracted information about cost and effects for both a short and long horizon. In each case, we calculated the incremental cost-effectiveness ratios (ICERs) using our abstractions and then converted our estimates into $US, year 2009 values.

1. Literature Review

1.1 Study Selection

Using published recommendations[10] for identification of cost-effectiveness studies, we searched MEDLINE (January 1997 to October 2009), EMBASE (January 1997 to June 2009) and the UK NHS Economic Evaluation Database[11] (1997 to October 2009) [for search strategies, see Appendix 1 in the Supplemental Digital Content 1, http://links.adisonline.com/PCZ/A73]). We also searched the bibliographies of included studies.

We included studies that evaluated the cost effectiveness of pharmacological agents in patients undergoing THR or TKR. Specifically, we focused our search on recent (1997 to October 2009) studies published in English that contained complete documentation of methods (as compared with abstracts or brief reports), had discrete information available for TKR or THR (i.e. not combined with other orthopaedic surgeries) and contained enough information to calculate an ICER for at least one of five important comparisons: (i) anticoagulants (fondaparinux, LMWH, warfarin) versus aspirin; (ii) LMWH versus warfarin; (iii) fondaparinux versus LMWH; (iv) comparisons including new oral anticoagulants; and (v) extended-duration prophylaxis (≥3 weeks with any agent) versus short-duration prophylaxis (<3 weeks with any agent). We did not analyse information about regimens not routinely recommended as sole therapy by the ACCP or AAOS. These include unfractionated heparin, parenteral thrombin inhibitors or non-pharmacological means such as intermittent pneumatic compression or graduated stockings. Two authors (AK and NR) evaluated each study for inclusion. Disagreements were resolved by discussion.

1.2 Study Abstraction and Quality Assessment

We derived an abstraction instrument based on the recommendations of the Panel on Cost-Effectiveness in Health and Medicine.[12-14] Two abstractors (NR and WC) assisted the primary author (AK) in recording, in duplicate, the description of the study setting, cohort age, economic perspective and presence of pharmaceutical industry sponsorship.

To summarize the cost-effectiveness information of our five main comparisons, we abstracted data on the incremental cost and effectiveness for both a short and a longer horizon when available. The horizon represents the period of time over which costs and effectiveness are aggregated. For certain diseases, such as the common cold, a short-horizon analysis may suffice. In other cases, long-term consequences must be accounted for, even for short-term interventions.[12-14] For the short horizon, we abstracted data on the projected costs incurred and VTE events avoided for the period closest to 90 days from surgery. For the purpose of calculating effectiveness, we

abstracted data on the combined incidence of DVT and PE that would be detected in routine clinical practice. If a study did not report such an outcome, we also accepted the incidence of radiographically detected events and noted the distinction. If effectiveness was defined only by life-years or QALYs, we recorded that information.

For the long horizon, we accepted any information that projected the cost and effectiveness for ≥1 year. We abstracted effectiveness information by prioritizing abstraction of the outcome of QALYs or unadjusted life-years.

For each study with missing information about drug regimen, dosage, duration of therapy, horizon of analysis, major bleeding rate, DVT, PE and death rate, we contacted corresponding authors first by email and then by letter. If the authors did not respond, we recorded the information as 'not reported'.

We adjusted all cost information to $US, year 2009 values, by inflating or deflating to the year 2005 according to readily available consumer price indices for each country,[15,16] converting to $US via WHO purchasing power parity indices,[17] and then inflating to $US using the Bureau of Labor Statistics consumer price calculator.[18] This approach followed the example of Bachmann.[19]

1.3 Study Quality

To assess study quality, we created an instrument adapted from 'Drummond's list'[20] and one other instrument from Brauer et al.[21] These included items about the use of cost data from a randomized controlled trial or other primary source, use of efficacy data from pooled results of a systematic review, identification of credible sources for all input parameters, appropriate calculation of an ICER and use of comprehensive one-way sensitivity analyses. The ICER is an expression of how much additionally it costs (in dollars) to achieve an additional unit of benefit (e.g. 1 more QALY). Policy makers are interested in the ICER value because it facilitates determination about whether newer, more effective interventions represent good value compared with existing, less expensive programmes.[22] The threshold for adoption in the US is thought to be somewhere between $US20 000 and $US100 000 per QALY gained.[23]

We also recorded quality items specific to VTE (including assessment of joint function following haemarthrosis, propagation of asymptomatic DVT to symptomatic PE, incidence of post-thrombotic syndrome, costs of major and minor bleeding) and future costs related to VTE (including blood monitoring and physician visits). Studies ignoring downstream bleeding consequences could make newer, more potent regimens appear more cost effective, whereas studies ignoring downstream costs of treating VTE will bias our interpretation in the other direction. We did not specifically document if individual studies included death costs related to VTE or bleeding. On the whole, death events were rare and the associated costs would be largely paid by the family of the patient and not the institution or health system, which was the economic perspective chosen by all but three of the studies analysed.

We did not pool the results of individual studies given the various modelling assumptions adopted by each author. Instead, we qualitatively compared studies to determine trends in the cost effectiveness of certain regimens compared with others.

1.4 Search Results

We identified 370 titles and abstracts meeting our search criteria. Of these, 56 were relevant and were entered for full-text review. Of these, 33 studies met all inclusion criteria[6-9,24-52] (figure 1).

Most studies were set in the US (14 of 33)[27-29,31-37,40,41,43,45] or Europe (14 of 33).[6-8,25,26,30,38,44,46-49,51,52] Twenty studies[6,7,24,27-37,40,41,43-46] adopted an institutional perspective; only three[9,48,49] adopted a societal perspective. Ten studies[6,8,24,28,31,36,46,47,50,52] reported pharmaceutical company sponsorship.

There was substantial variation in the quality of reporting. Only six of the 33 studies reported performing a systematic review and meta-analysis of

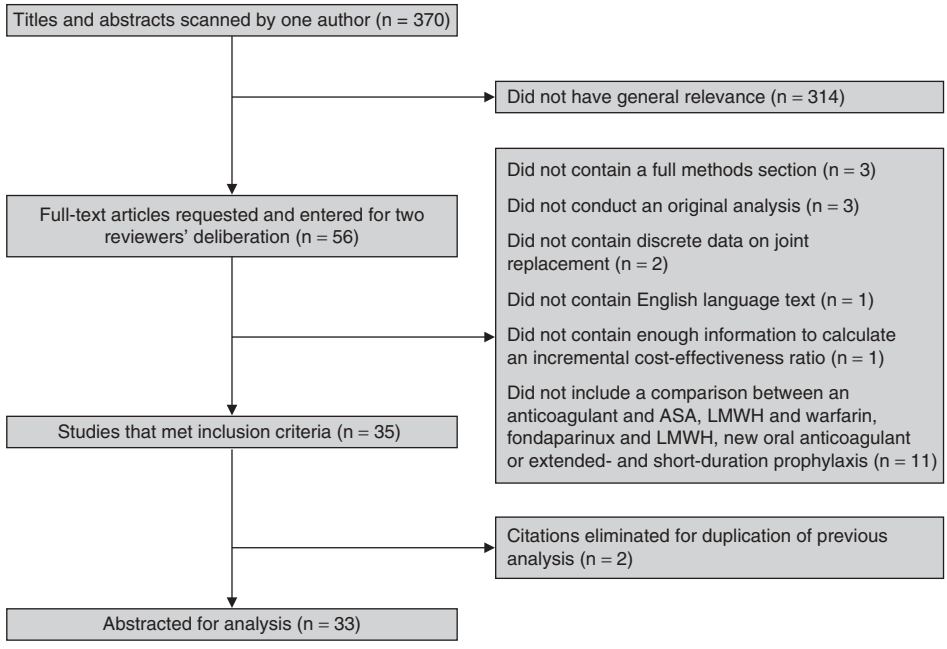


Fig. 1. Flow diagram of article selection. **ASA** = acetylsalicylic acid (aspirin); **LMWH** = low-molecular-weight heparin.

efficacy data.[9,26,28,35,36,50] In addition, only 13 studies[6,9,25,26,28,32-35,44,47,48,52] documented comprehensive use of one-way sensitivity analysis. Only four of 33 studies[9,28,49,50] measured effectiveness in QALYs to at least the 1-year horizon (table I, and table A1 in the Supplemental Digital Content).

2. Anticoagulants versus Aspirin

We included two studies[24,25] with three comparisons of an anticoagulant with aspirin (table II). In all three comparisons, results were available for THR exclusively. Sarasin and Bounameaux[25] found that the ICER was $US1700 per VTE avoided for 4 weeks of warfarin compared with aspirin and $US1300 per VTE avoided for 4 weeks of LMWH compared with aspirin. There was no apparent pharmaceutical company sponsorship for that study. The final comparison, sponsored by sanofi-aventis, the manufacturer of enoxaparin, was set in South Africa and reported an ICER of $US7200 per VTE avoided for 10 days of enoxaparin compared with 10 days of aspirin.

3. Low-Molecular-Weight Heparin (LMWH) versus Warfarin

We included 15 studies with comparisons of LMWH and warfarin[9,25-37,50] (table III). Twelve compared these agents in patients receiving THR. Of those documenting a short-horizon cost-effectiveness result, three[25,34,50] found that the ICER for LMWH was ≤$US2000 per VTE avoided compared with warfarin. In two other studies,[28,29] LMWH cost an additional $US2100 per VTE avoided. In a sixth study,[32] LMWH cost $US5200 per VTE avoided. In the next study,[9] LMWH cost $US109 000 per VTE avoided. This study by Skedgel et al.[9] examined 4 additional weeks (in addition to the hospital period) of LMWH compared with 4 additional weeks of warfarin. It found that the cost, in Canada, would be almost 10-fold higher for LMWH given the significant proportion of patients (39% at

Table I. Descriptive characteristics of studies included in the systematic review of venous thromboembolism (VTE) pharmacological prophylaxis after total hip and knee replacement

Characteristic	Studies [n (%)[a]]
Setting	
USA	14 (42)
Canada	4 (12)
Europe	14 (42)
South Africa	1 (3)
Economic perspective	
Institutional	20 (61)
National health system	10 (30)
Societal	3 (9)
Sponsorship[b]	
Pharmaceutical sponsor	10 (30)
Pharmaceutical grant	9 (27)
Pharmaceutical consultants	2 (6)
Government agency	3 (9)
None reported	9 (27)
Comparison type	
Anticoagulant vs aspirin	2 (6)
LMWH vs warfarin	15 (45)
Fondaparinux vs LMWH	10 (30)
Comparisons with new oral anticoagulants	2 (6)
Extended- vs short-duration prophylaxis	9 (27)
Quality inventory[c]	
Costs measured through primary source?	16 (48)
Effectiveness calculated using pooled results of systematic review?	6 (18)
Data sources comprehensively documented and credible?	29 (88)
Costs and effects discounted (for studies with horizon 1 y or more)?	6 (86)[d]
ICER calculated correctly?	30 (91)
One-way sensitivity analysis used comprehensively?	13 (39)
Other distinguishing features	
Effectiveness measured in QALYs at a horizon of at least 1 y?	7 (21)
Asymptomatic VTE adequately addressed?	18 (54)
Post-thrombotic syndrome adequately addressed?	10 (30)
Major bleeding included in cost calculation?	28 (85)

a Of 33 studies.

b If both pharmaceutical and government sponsorship, pharmaceutical sponsorship was recorded.

c Derived from quality scales published separately by Drummond et al.[20] and Brauer et al.[21]

d Of a total of seven studies.

ICER = incremental cost-effectiveness ratio; **LMWH** = low-molecular-weight heparin.

baseline) that would require *daily* nursing supervision of LMWH injection in their homes than the same proportion that would require *weekly* home phlebotomy for monitoring INR while using warfarin. In the remaining four studies of short horizon,[26,30,31,36] warfarin dominated LMWH.

In two studies with a long horizon, results conflicted, with one study[28] finding that LMWH dominated warfarin while the other[37] found the opposite.

In comparisons that analysed cost effectiveness in the setting of TKR (or TKR cases combined with THR cases), LMWH dominated or cost <$US2000 per VTE avoided in five studies.[27,33-35,50] In the final study,[26] warfarin dominated LMWH.

Eight of 15 studies comparing LMWH with warfarin reported some pharmaceutical company sponsorship, grant support or involvement of pharmaceutical company consultants. In each case, the pharmaceutical company was the manufacturer of LMWH; either sanofi-aventis, Pfizer or a company that merged with these two. All but two[31,36] of these eight found favourable cost-effectiveness ratios for LMWH. The two studies[9,26] by government agencies indicated that LMWH was either poor value for its cost or was dominated by warfarin.

4. Fondaparinux versus LMWH

We included ten studies with comparisons of fondaparinux and LMWH[7,8,38-45] (table IV). Nine of ten analysed prophylaxis for THR. Six studies[7,8,38,39,41,45] analysed cost effectiveness over a short horizon. In all six, fondaparinux dominated or cost ≤$US1300 per VTE avoided. In four studies with a long horizon, fondaparinux dominated LMWH. In a fifth, LMWH cost $US40 per VTE avoided.

Of the eight studies reporting cost-effectiveness results for TKR,[7,8,38,39,42-45] all but one found that fondaparinux dominated LMWH over the short and long horizon. In this study,[42] fondaparinux cost an additional $US660 per VTE avoided.

Among the ten studies comparing fondaparinux with LMWH, a pharmaceutical company sponsored one and supported five more through

Table II. Summary of cost-effectiveness studies comparing anticoagulants with acetylsalicylic acid (ASA; aspirin)

Study	Comparator 1	ASA regimen	Horizon (days)	Major bleeding rate (%)[a]	DVT rate (%)[a]	PE rate (%)[a]	Death rate (%)[a]	Cost-effectiveness result reported by study authors ($US)[c]
Total hip replacement results from short-horizon analysis								
Sarasin and Bounameaux[25]	Enoxaparin 40 mg od×10 d	300 mg od×10 d	90	0.49/0.48	5.59/8.79[b]	1.12/1.76	NR/NR	1300 per VTE avoided
	Warfarin dose NR×28 d post-discharge	160 mg od×28 d post-discharge	90	0.59/0.48	6.52/8.79[b]	1.30/1.76	NR/NR	1700 per VTE avoided
Abdool-Carrim et al.[24]	Enoxaparin 40 mg or dalteparin 4000–5000 IU od×28 d post-discharge	160 mg od×28 d post-discharge	NR	2.00/0.70	21.0/35.0[b]	NR/NR	1.10/1.50	7200 per VTE avoided

a Figures presented as anticoagulant/ASA.

b DVT rate not specified according to clinical vs radiographic and proximal vs distal.

c Cost-effectiveness result is the ICER. To arrive at ICER values, incremental costs reported in foreign currencies were inflated or deflated according to readily available consumer price indices, converted to $US via 2005 WHO purchasing price parity indices, and then inflated to year 2009 values using the Bureau of Labor Statistics consumer price calculator.[18]

DVT = deep venous thrombosis; **ICER** = incremental cost-effectiveness ratio; **NR** = not reported; **od** = once daily; **PE** = pulmonary embolism; **VTE** = venous thromboembolism.

grants. In each case the sponsor or grantor was sanofi-aventis, the manufacturer of enoxaparin (the inferior comparator). Each result demonstrated good value with dominance by the use of fondaparinux.

5. Comparisons Including New Oral Anticoagulants

Only two studies to date have made comparisons with new oral anticoagulants (see table V). In the only one that made this comparison in patients undergoing THR, Wolowacz et al.[52] found that dabigatran dominated LMWH over a 60-year horizon (equivalent to a lifetime analysis, given the elderly age of the average patient undergoing THR).

In the setting of TKR, McCullagh et al.,[51] found that, in the short horizon of 180 days, rivarobaxan dominated both LMWH and dabigatran; dabigatran cost an additional $US750 per VTE avoided compared with LMWH. In the long horizon, Wolowacz et al.[52] found that dabigatran dominated LMWH.

The study by Wolowacz et al.[52] was sponsored by Boehringer Ingelheim, the manufacturer of dabigatran, whereas McCullagh et al.[51] reported no sponsorship or support.

6. Extended-Duration versus Short-Duration Prophylaxis

We found nine studies[6,9,30,46-51] with comparisons of extended- versus short-duration prophylaxis in patients undergoing THR (table VI). Among short-horizon results, three studies[30,46,51] with five comparisons found that extended-duration therapy after THR either dominated short-duration prophylaxis or the ICER was <$US120 per VTE avoided. In Skedgel et al.,[9] extended-duration warfarin prophylaxis cost an additional $US3200 per VTE avoided but extended-duration LMWH cost an additional $US27 400 per VTE avoided. In five other studies, the ICER for extended-duration therapy was between $US7800 and $US13 200 per VTE avoided. In McCullagh et al.,[51] dabigatran administered for 35 days cost $US730 000 per VTE

Table III. Summary of cost-effectiveness studies comparing low-molecular-weight heparin (LMWH) with warfarin

Study	Horizon	LMWH regimen	Warfarin regimen	Major bleeding rate (%)[a]	DVT rate (%)[a]	PE rate (%)[a]	Death rate (%)[a]	Cost-effectiveness result ($US)[c]
Total hip replacement results from short-horizon analysis								
Sarasin and Bounameaux[25]	90 d	Enoxaparin 40 mg + tinzaparin 4000–5000 IU od × 28 d post-discharge	INR 2–3 × 28 d post-discharge	0.49/0.59	5.59/6.52[d]	1.12/1.30	NR/NR	312 per VTE avoided
Hull et al.[34]	NR	Tinzaparin 5250 IU od × 9 d	INR 2–3 × 9 d	2.80/1.50	20.8/23.2[b]	NR/NR	NR/NR	970 per VTE avoided
Dranitsaris et al.[50]	10 d	Dalteparin 5000 IU od × 10 d	Warfarin INR 2–3 × 10 d	6.6/4.5	4.4/6.7	0/0	0/0	1300 per VTE avoided
Botteman et al.[28]	NR	Enoxaparin 30 mg bid × 7 d	5 mg od × 7 d	NR/NR	13.6/21.3[d]	NR/NR	0.70/1.10	2100 per VTE avoided
Caprini et al.[29]	7 d	Enoxaparin 30 mg bid × 7 d	INR 2–3 × 10 d	1.20/0.50	2.00/4.00	0.50/0.90	NR/NR	2100 per VTE avoided
Friedman and Dunsworth[32]	21 d	Enoxaparin 40 mg od × 21 d	5 mg od × 21 d	0/1.00	8.00/10.0	0/0.99	NR/NR	5200 per VTE avoided
Skedgel et al.[9]	90 d	Enoxaparin 40 mg od × 28 d post-discharge	5 mg od × 28 d post-discharge	0.11/0.54	1.10/1.57	0.24/0.34	0.03/0.05	109 000 per VTE avoided[e]
Dahl and Pleil[30]	35 d	Dalteparin 5000 IU od × 7–15 d	5 mg od × 7–15 d	NR/NR	8.50/8.30	2.30/0.90	NR/NR	Warfarin dominates
Anderson et al.[26]	90 d	Enoxaparin 30 mg bid × 7–14 d	5 mg od; duration NR	2.04/0.98	2.40/2.40	1.10/1.10	0.10/0.09	Warfarin dominates
Francis et al.[31]	7 d	Enoxaparin (dose and duration NR)	NR	1.70/1.00	8.30/5.30	0.50/0	NR/NR	Warfarin dominates
Wade and Hawkins[36]	30 d	Enoxaparin 40 mg od × 30 d	5 mg od × 30 d	0/0	6.20/2.90	0/0	NR/NR	Warfarin dominates
Total hip replacement results from long-horizon analysis								
Botteman et al.[28]	Lifetime	Enoxaparin 30 mg bid × 7 d	5 mg od × 7 d	NR/NR	NR/NR	NR/NR	NR/NR	LMWH dominates
Wade and Spruill[37]	1 y	Enoxaparin 30 mg bid × 7 d	5 mg od × 7 d	2.80/1.30	4.75/3.45[d]	1.95/0.75	NR/NR	Warfarin dominates
Total knee replacement results from short-horizon analysis								
Nerurkar et al.[35]	Hospital period	Enoxaparin × 5–6 d, dose NR	5–6 d, dose NR	2.00/3.00	5.85/3.55[b]	NR/NR	1.22/1.92	LMWH dominates
Hull et al.[34]	NR	Tinzaparin 5250 IU od × 9 d	INR 2–3 × 9 d	2.80/0.90	45.0/54.9[b]	NR/NR	NR/NR	950 per VTE avoided
Hawkins et al.[33]	NR	Enoxaparin 30 mg bid × 4 d	5 mg od × 4 d	2.10/1.80	20.90/35.1[b]	NR/NR	0/0	1100 per VTE avoided

Continued next page

Table III. Contd

Study	Horizon	LMWH regimen	Warfarin regimen	Major bleeding rate (%)[a]	DVT rate (%)[a]	PE rate (%)[a]	Death rate (%)[a]	Cost-effectiveness result ($US)[c]
Dranitsaris et al.[50]	35 d	Dalteparin 5000 IU od×10 d	Warfarin INR 2–3×10 d	6.6/4.5	4.4/5.8	0/0	0/0	1400 per VTE avoided
Anderson et al.[26]	90 d	Enoxaparin 30 mg bid×7–14 d	5 mg od; duration NR	2.04/0.98	1.92/1.92	0.87/0.87	0.23/0.22	Warfarin dominates
Total hip replacement combined with total knee replacement result								
Bell and Goldhaber[27]	6 mo	Ardeparin (dose and duration NR)	NR	NR/NR	NR/NR	NR/NR	NR/NR	LMWH dominates

a Figures presented as LMWH/warfarin.

b Radiographic VTE rate.

c Cost-effectiveness result is the ICER. To arrive at ICER values, incremental costs reported in foreign currencies were inflated or deflated according to readily available consumer price indices, converted to $US via 2005 WHO purchasing price parity indices, and then inflated to 2009 values using the Bureau of Labor Statistics consumer price calculator.[18]

d DVT rate (i.e. symptomatic vs radiographic) not specified.

e ICER compares extended LMWH with extended-duration warfarin.

bid = twice daily; **DVT** = deep venous thrombosis; **ICER** = incremental cost-effectiveness ratio; **INR** = international normalized ratio; **NR** = not reported; **od** = once daily; **PE** = pulmonary embolism; **VTE** = venous thromboembolism.

avoided compared with short-duration LMWH; the high ICER results mainly from the many-fold increased bleeding rates found with dabigatran than with LMWH (2.0% vs 0.08%).

Among two THR studies with long-horizon results available, Bischof et al.[6] found that extended-duration fondaparinux dominated short-duration fondaparinux. Haentjens et al.[49] found that extended-duration enoxaparin cost an additional $US9300 per QALY gained compared with short-duration enoxaparin.

We found only two studies for TKR. At a 35-day horizon, Dranitsaris et al.[50] found that the extended-duration dalteparin cost an additional $US14 600 per VTE compared with short-duration warfarin and $US60 000 per VTE compared with short-duration dalteparin. At a 1-year horizon, Haentjens et al.[49] found that extended-duration enoxaparin cost an additional $US73 300 per QALY gained compared with short-duration enoxaparin.

Six of the nine studies comparing extended- with short-duration therapy included pharmaceutical company sponsorship or grant support. There was no clear trend among the results with respect to the presence of sponsorship, although two of the three studies sponsored exclusively by a government agency found that extended-duration therapy with LMWH or dabigatran delivered improved effectiveness at a relatively high cost ($US27 400–730 000 per VTE avoided). As mentioned above, the third study by Haentjens et al.[49] found that extended-duration LMWH was clearly cost effective after THR but poorer value after TKR.

7. Discussion

Although multiple VTE prophylaxis regimens are supported by the ACCP and the AAOS, our systematic review suggests that not all of them may be cost effective relative to other regimens. There was no consensus about the cost effectiveness of LMWH compared with warfarin. In contrast, fondaparinux dominated LMWH in nearly every comparison we found. Extended-duration prophylaxis with LMWH after THR appeared to be cost effective, with multiple

Table IV. Summary of cost-effectiveness studies comparing fondaparinux with low-molecular-weight heparin (LMWH)

Study	Horizon	Fondaparinux regimen	LMWH regimen	Major bleeding (%) rate[a]	DVT rate (%)[a]	PE rate (%)[a]	Death rate (%)[a]	Cost-effectiveness result reported by study authors ($US)[b]
Total hip replacement results from short-horizon analysis								
Dranitsaris et al.[39]	90 d	Dose NR; 7 d	Dose NR; 7 d	Difference = 0	Fondaparinux 0.69% more averted	Difference = 0.41	NR/NR	Fondaparinux dominates
Bjorvatn and Kristiansen[8]	90 d	2.5 mg od × 7 d	40 mg od × 7 d	NR/NR	1.84/2.71	0.58/1.09	0.14/0.22	Fondaparinux dominates
Spruill et al.[41]	10 d	3 mg od; duration NR	30 mg bid; duration NR	4.5/3.5	0.90/2.90	NR/NR	0/0	Fondaparinux dominates
Sullivan et al.[45]	90 d	2.5 mg od × 7 d	30 mg bid × 7 d	NR/NR	Fondaparinux 1.16% more VTEs averted	NR/NR	NR/NR	Fondaparinux dominates
Wade et al.[40]a	11 d	2.5 mg od × 5–9 d	30 mg bid × 5–9 d	2.88/2.71	2.02/3.01	0.58/1.09	0.10/0.18	Fondaparinux dominates
Annemans et al.[42]	90 d	Dose NR; 7 d	Dose NR; 7 d	2.87/2.70	1.85/2.73	0.58/1.09	0.10/0.18	1300 per VTE avoided
Total hip replacement results from long-horizon analysis								
Annemans et al.[42]	5 y	Dose NR; 7 d	Dose NR; 7 d	2.88/2.71	2.02/3.01	0.58/1.09	0.10/0.18	Fondaparinux dominates
Gordois et al.[38]	5 y	Dose NR; 7 d	40 mg od × 7 d	2.80/2.60	Fondaparinux 1.50% more total VTEs averted	0.58/1.09	Difference = 0.8	Fondaparinux dominates
Sullivan et al.[45]	5 y	2.5 mg od × 7 d	30 mg bid × 7 d	NR/NR	Fondaparinux 1.16% more VTEs averted	NR/NR	NR/NR	Fondaparinux dominates
Szucs et al.[7]	5 y	Dose NR; 7 d	Dose NR; 7 d	2.85/2.69	1.96/2.88	0.59/1.09	0.11/0.18	Fondaparinux dominates
Lundkvist et al.[44]	5 y	Dose NR; 7 d	40 mg od × 7 d	NR/NR	1.84/2.71	0.58/1.09	0.11/0.19	40 per VTE avoided
Total knee replacement results from short-horizon analysis								
Bjorvatn and Kristiansen[8]	90 d	2.5 mg od × 7 d	40 mg od × 7 d	NR/NR	1.49/2.73	0.66/1.19	0.18/0.35	Fondaparinux dominates
Dranitsaris et al.[39]	90 d	Dose NR; 7 d	Dose NR; 7 d	Difference = 0%	Fondaparinux 1.27% more averted	Difference = 0.54	NR/NR	Fondaparinux dominates
Spruill et al.[43]	NR	2.5 mg od × 4–5 d	30 mg bid × 4–5 d	2.1/0.20	2.40/5.40	0.20/0.80	0/0	Fondaparinux dominates
Sullivan et al.[45]	90 d	2.5 mg od × 7 d	30 mg bid × 7 d	NR/NR	Fondaparinux 1.78% more VTEs averted	NR/NR	NR/NR	Fondaparinux dominates
Annemans et al.[42]	90 d	Dose NR; 7 d	Dose NR; 7 d	2.87/2.71	1.50/2.75	0.66/1.19	0.12/0.19	660 per VTE avoided

Continued next page

Table IV. Contd

Study	Horizon	Fondaparinux regimen	LMWH regimen	Major bleeding (%) rate[a]	DVT rate (%)[a]	PE rate (%)[a]	Death rate (%)[a]	Cost-effectiveness result reported by study authors ($US)[b]
Total knee replacement results from long-horizon analysis								
Annemans et al.[42]	5 y	Dose NR; 7 d	Dose NR; 7 d	2.87/2.71	1.68/3.11	0.66/1.19	0.12/0.19	Fondaparinux dominates
Gordois et al.[38]	5 y	2.5 mg od×7 d	40 mg od×7 d	2.8/2.6	Fondaparinux 1.95% more total VTEs averted	0.66/1.19	Difference = 0.7	Fondaparinux dominates
Lundkvist et al.[44]	5 y	2.5 mg od×7 d	40 mg od×7 d	NR/NR	1.49/2.73	0.66/1.19	0.12/0.20	Fondaparinux dominates
Sullivan et al.[45]	5 y	2.5 mg od×7 d	30 mg bid×7 d	NR/NR	Fondaparinux 1.78% more VTEs averted	NR/NR	NR/NR	Fondaparinux dominates
Szucs et al.[7]	5 y	Dose NR; 7 d	Dose NR×8 d	2.85/2.69	1.60/2.92	0.65/1.20	0.11/0.19	Fondaparinux dominates

a Figures presented as fondaparinux/LMWH.

b Cost-effectiveness result is the ICER. To arrive at ICER values, incremental costs reported in foreign currencies were inflated or deflated according to readily available consumer price indices, converted to $US via 2005 WHO purchasing price parity indices, and then inflated to 2009 values using the Bureau of Labor Statistics consumer price calculator.[18]

bid = twice daily; **DVT** = deep venous thrombosis; **ICER** = incremental cost-effectiveness ratio; **NR** = not reported; **od** = once daily; **PE** = pulmonary embolism; **VTE** = venous thromboembolism.

studies indicating that extended-duration prophylaxis dominated short-duration LMWH or cost no more than an additional $US10 000 per VTE avoided. The small numbers of studies, predominance of studies analysing only a short horizon, lack of established cost-effectiveness thresholds for VTE-based effectiveness units and reliance by study authors on venographic endpoints prohibit robust conclusions about the comparisons analysed.

Comparisons of our work with previous reviews of the economic literature are limited by differences in type of surgery included and publication dates of the included articles. Sullivan et al.[53] summarized the prophylaxis literature between 1984 and 2000 and found that most studies presented consistent findings, including that LMWH is cost effective compared with warfarin. Our results do not support this conclusion. Sullivan et al.[53] based their conclusions on many studies that we excluded because they were published prior to 1997 or that included outcomes from patients undergoing hip fracture surgery. We believe temporal trends[54,55] in the care of THR and TKR necessitated excluding earlier studies. We also felt that hip fracture surgery identified a distinctive patient population with respect to cost, risk and benefit issues.[4] Similar to our findings, Sullivan et al.[53] also found that extended-duration LMWH was generally cost effective compared with short-duration therapy.

Ivanovic et al.[56] summarized the literature on fondaparinux. These authors concluded that fondaparinux was *more* cost effective than LMWH (enoxaparin) 40 mg daily initiated *preoperatively* but *less* cost effective than LMWH 30 mg twice daily initiated *postoperatively*. Our review did not specifically compare the cost effectiveness of regimens with LMWH initiated at different times but we found that fondaparinux dominated LMWH in all but one when considering the longer horizon. LMWH dosages in the included studies were evenly distributed between 40 mg daily and 30 mg twice daily. Ivanovic et al.[56] also reported not being able to calculate ICERs for two studies, whereas we were able to calculate them based on data presented in tables included by the study authors.

Table V. Summary of cost-effectiveness studies of new oral anticoagulants[a]

Study	Exact horizon	Oral anticoagulant	Comparator	Major bleeding rate (%)[b]	DVT rate (%)[b]	PE rate (%)[b]	Death rate (%)[b]	Cost-effectiveness result ($US)[c]
Total hip replacement result from long-horizon analysis								
Wolowacz et al.[52]	60 y	Dabigatran 220 mcg × 28–35 d	Enoxaparin 40 mg × 28–35 d	2.0/1.6	4.6/4.8	0.9/0.9	0.4/0.4	Dabigatran dominates
Total knee replacement results from short-horizon analysis								
McCullagh et al.[51]	180 d	Rivarobaxan 10 mg × 14 d	Enoxaparin 40 mg × 10 d	0.57/0.08	0.87/1.8	0/0.10	0/0	Rivaroxaban dominates
			Dabigatran 220 mcg × 10 d	0.57/1.5	0.87/2.7	0/0	0/0	Rivaroxaban dominates
		Dabigatran 220 mcg × 10 d	Enoxaparin 40 mg × 10 d	1.5/1.3	2.7/1.8	0/0.10	0/0	750 per VTE avoided
Total knee replacement result from long-horizon analysis								
Wolowacz et al.[52]	60 y	Dabigatran 220 mcg × 6–10 d	Enoxaparin 40 mg × 6–10 d	1.5/1.3	12.1/12.4	2.1/2.2	1.7/1.7	Dabigatran dominates

a All doses are once daily.

b Figures presented as new oral anticoagulant/LMWH.

c Cost-effectiveness result is the ICER. To arrive at ICER values, incremental costs reported in foreign currencies were inflated or deflated according to readily available consumer price indices, converted to $US via 2005 WHO purchasing price parity indices, and then inflated to 2009 values using the Bureau of Labor Statistics consumer price calculator.[18]

DVT = deep venous thrombosis; **ICER** = incremental cost-effectiveness ratio; **LMWH** = low-molecular-weight heparin; **PE** = pulmonary embolism; **VTE** = venous thromboembolism.

Wolowacz et al.[57] also published a review discussing the evolution of model building over a 20-year time span (1987–2006). In terms of quality, the findings of that review were generally consistent with the abstractions we performed, particularly with respect to the paucity of studies measuring QALYs over a sufficiently long period. Unlike their review, we abstracted cost and effect information and independently calculated ICERs for each comparison discussed. We converted costs to $US, year 2009 values and measured effects in common units (total VTE events avoided for short-horizon studies and QALYs for long-horizon studies). This facilitated comparisons between the multiple regimens supported by major professional societies.

The most salient finding of our review is that fondaparinux dominates LMWH; however, these results should be interpreted cautiously. There have been only four randomized controlled trials comparing fondaparinux with enoxaparin[58-61] and only one[58] involved patients with TKR surgery. A summary estimate of risk calculated by Turpie et al.[62] suggested that fondaparinux offers a 55% reduction in the odds of venographic VTE but no difference in the incidence of symptomatic VTE at postoperative day 11 when screening venography was performed. The studies of cost effectiveness evaluating fondaparinux generally extrapolated these short-horizon venographic rates to estimate the number of symptomatic VTE events. Recent evidence[63] suggests that the ratio of asymptomatic venographic DVT rate to symptomatic DVT rate is between 3 and 7 for THR and between 15 and 24 for TKR. However, these ratios came from trials using enoxaparin only. Although they do not address this point specifically for fondaparinux, the 2008 ACCP guidelines[4] state that initial efficacy studies using venographic endpoints should be followed with trials that use symptomatic (and objectively confirmed) VTE as endpoints.

There is less conclusive evidence about the duration of prophylaxis, although extended prophylaxis with LMWH appears cost effective compared with short-duration therapy in the case of THR surgery. Authors of cost-effectiveness studies included in this review generally summarized efficacy of extended-duration prophylaxis with LMWH using one or more of the seven

Table VI. Summary of cost-effectiveness studies of extended- versus short-duration therapy

Study	Exact horizon	Extended duration[a]	Short duration[a]	Major bleeding rate (%)[b]	DVT rate (%)[b]	PE rate (%)[b]	Death rate (%)[b]	Cost-effectiveness result ($US)[c]
Total hip replacement results from short-horizon analysis								
Bergqvist[46]	19–23 d post-discharge	Enoxaparin 40 mg×30 d	Enoxaparin 40 mg×9 d	2.04/4.17	1.53/5.34	0/1.53	0/0	Extended-duration LMWH dominates
Dahl and Pleil[30]	35 d	Dalteparin 5000 IU ×28–35 d	Dalteparin 5000 IU ×7–15 d	NR/NR	5.50/8.5	0.5/2.3	NR/NR	Extended-duration LMWH dominates
			Warfarin	NR/NR	5.5/8.3	0.5/0.9	NR/NR	120 per VTE avoided
McCullagh et al.[51]	180 d	Rivarobaxan 10 mg×35 d	Dabigatran 220 mcg×14 d	0.08/2.0	0.29/0.93	0.40/0.50	0/0	Extended-duration rivaroxaban dominates
			Enoxaparin 40 mg×14 d	0.08/0.08	0.29/2.2	0.12/0.50	0/0	Extended-duration rivaroxaban dominates
		Dabigatran 220 mcg×35 d	Enoxaparin 40 mg×14 d	2/0.08	0.93/2.2	0.40/0.50	0/0	730 000 per VTE avoided
Skedgel et al.[9]	90 d	Warfarin 5 mg×28 d post-discharge	Regimen for hospital period NR	0.54/0.11	1.57/3.28	0.29/0.61	0.05/0.10	3200 per VTE avoided
		LMWH 40 mg×28 d post-discharge		0.11/0.11	1.10/3.28	0.20/0.61	0.03/0.10	27 400 per VTE avoided
Davies et al.[47]	90 d	Enoxaparin 40 mg×hospitalization period+21 d	Enoxaparin 40 mg for hospitalization period	NR/NR	1.8/7.4	NR/NR	0.1/0.7	7800 per VTE avoided
Dranitsaris et al.[50]	35 d	Dalteparin 5000 IU ×35 d	Warfarin INR 2–3×10 d	6.6/4.5	3.72/6.7	0/0	0/0	8000 per VTE avoided
			Dalteparin 5000 IU ×10 d	6.6/6.7	3.72/5.3	0/0	0/0	13 200 per VTE avoided
Detournay et al.[48]	30–35 d	Enoxaparin 40 mg×30–35 d	Enoxaparin 40 mg ×7–14 d	NR/NR	Extended 16.0–21.1% more events avoided	NR/NR	Extended 0.60–0.78% more events avoided	10 000 per VTE avoided
Bischof et al.[6]	30 d	Fondaparinux×28 d; dose NR	Fondaparinux×7 d; dose NR	NR/NR	Fondaparinux 1.6% more events avoided	Fondaparinux 0.5% more events avoided	0/0.1	13 300 per life-year gained
Total hip replacement results from long-horizon analysis								
Bischof et al.[6]	5 y	Fondaparinux×28 d; dose NR	Fondaparinux×7 d; dose NR	NR/NR	NR/NR	NR/NR	NR/NR	Extended-duration fondaparinux dominates

Continued next page

Table VI. Contd

Study	Exact horizon	Extended duration[a]	Short duration[a]	Major bleeding rate (%)[b]	DVT rate (%)[b]	PE rate (%)[b]	Death rate (%)[b]	Cost-effectiveness result ($US)[c]
Haentjens et al.[49]	1 y	Enoxaparin×42 d; dose NR	Enoxaparin×12 d; dose NR	1.7/1.7	5.12/8.95	NR/NR	NR/NR	9300 per QALY gained
Total knee replacement results from short-horizon analysis								
Dranitsaris et al.[50]	35 d	Dalteparin 5000 IU ×35 d	Warfarin INR 2–3×10 d	6.7/4.8	4.0/5.8	0/0	0/0	ICER 14 600 per VTE
			Dalteparin 5000 IU ×10 d	6.7/6.9	4.0/4.4	0/0	0/0	60 600 per VTE avoided
Total knee replacement result from long-horizon analysis								
Haentjens et al.[49]	1 y	Enoxaparin 40 mg×42 d	Enoxaparin×12 d; dose NR	0.5/0.5	6.81/7.70	NR/NR	NR/NR	73 300 per QALY gained

a All doses are once daily unless otherwise specified.

b Figures presented as extended/short duration.

c Cost-effectiveness result is the ICER. To arrive at ICER values, incremental costs reported in foreign currencies were inflated or deflated according to readily available consumer price indices, converted to $US via 2005 WHO purchasing price parity indices, and then inflated to 2009 values using the Bureau of Labor Statistics consumer price calculator.[18]

DVT = deep venous thrombosis; **ICER** = incremental cost-effectiveness ratio; **INR** = international normalized ratio; **LMWH** = low-molecular-weight heparin; **NR** = not reported; **PE** = pulmonary embolism; **VTE** = venous thromboembolism.

randomized controlled trials[64-70] that reported on the efficacy of extended-duration prophylaxis. At least two of these trials[64,65] did not require venography at the time of discharge from the hospital, permitting assessment of symptomatic VTE rates from 4 to 7 weeks after operation. We cannot draw firm conclusions on the question of extended duration versus short duration of therapy with other agents that have not been studied extensively. Our review also suggests that there is insufficient cost-effectiveness evidence to support extended prophylaxis for TKR. The most recent update of the ACCP guidelines ''recommends'' extended prophylaxis for THR and ''suggests'' extended prophylaxis for TKR.

Limitations to our work include differences in economic perspective and setting. As our results overwhelmingly suggest that fondaparinux dominates LMWH, we believe our conclusions are sound for this comparison, keeping in mind the absence of trial data measuring symptomatic endpoints. The economic perspective did not appear to explain the variations in results found, but we did not have sufficient numbers of studies within each major comparison to make firm statements about the influence of individual differences in analytic methods. Although we converted from foreign currencies to $US using purchasing power parity, cost structures between countries may not be comparable.[71]

We also acknowledge the potential bias exerted by pharmaceutical company sponsorship of multiple studies. This bias could have played a role in the comparisons between LMWH and warfarin and extended- versus short-duration therapy. They do not appear to have played a role in the comparisons including fondaparinux. Multiple studies sponsored by the manufacturer of LMWH found fondaparinux to dominate LMWH. However, we generally did not have sufficient numbers of studies within each comparison type to determine if variation in study results was related to pharmaceutical company sponsorship.

Another major limitation is that there is no established threshold for declaring a prophylaxis regimen cost effective when disease-based units are used to express effectiveness. The QALY permits comparison of the value of interventions

across diseases given that the utilities that are used to calculate them are standardized to estimates between 0 and 1, where 1 represents perfect health and 0 represents death.

Another limitation includes absence of cost-effectiveness analyses about certain comparisons such as fondaparinux versus warfarin, fondaparinux versus aspirin, and low-intensity warfarin (INR < 2) versus any of the other regimens. We also acknowledge the possibility of English language and publication bias, as with any systematic review.

The demand for cost-effectiveness research is growing at a fervent pace. In early 2009, the US Government dedicated $US1.1 billion to comparative effectiveness research, including cost-effectiveness research.[72] The US Centers for Disease Control adopted the results of cost-effectiveness research when it prepared guidelines[73] about screening for HIV infection. Similarly, the US Preventive Services Task Force incorporated model results when it updated its most recent colorectal cancer screening recommendations.[74] As the demand for cost-effectiveness work grows, the need to be able to summarize and standardize the information will also grow. Our work was a comprehensive systematic review of the cost-effectiveness literature regarding VTE prophylaxis for patients undergoing total joint replacement. In addition, we improved upon previous reviews by standardizing cost-effectiveness information to a common currency and effectiveness unit.

8. Conclusions

We found that fondaparinux dominated LMWH in virtually all studies we analysed, but firm conclusions cannot be made until trial data are available that measure symptomatic VTE rates. Extended-duration LMWH prophylaxis also appears cost effective compared with short-duration prophylaxis in the case of THR. There is limited evidence to determine the cost effectiveness of other regimens, including extended-duration fondaparinux, extended-duration LMWH after TKR, prophylaxis with new oral anticoagulants, low-intensity warfarin therapy or aspirin. These knowledge gaps represent important areas for future research.

Acknowledgements

The authors have no conflicts of interest that are directly relevant to the content of this review. This work was performed at Boston Medical Center affiliated with Boston University School of Medicine.

References

1. Kapoor A, Labonte A, Winter M, et al. Venous thromboembolism after total hip and knee replacement in older adults with single and co-occurring comorbidities. Cleve Clin J Med 2010 Mar; 77 (Suppl. 1): eS8 [online]. Available from URL: http://www.ccjm.org/content/77/Electronic_Suppl_1/eS8.full.pdf+html [Accessed 2010 Mar 16]
2. Kurtz S, Ong K, Lau E, et al. Projections of primary and revision hip and knee arthroplasty in the United States from 2005 to 2030. J Bone Joint Surg Am 2007 April 1; 89 (4): 780-5
3. Zhan C, Kaczmarek R, Loyo-Berrios N, et al. Incidence and short-term outcomes of primary and revision hip replacement in the United States. J Bone Joint Surg Am 2007 Mar; 89 (3): 526-33
4. Geerts WH, Pineo GF, Heit JA, et al. Prevention of venous thromboembolism: the Eighth ACCP Conference on Antithrombotic and Thrombolytic Therapy. Chest 2008; 133 (3 Suppl.): 381S-453
5. American Academy of Orthopaedic Surgeons. American Academy of Orthopaedic Surgeons clinical guideline on prevention of pulmonary embolism in patients undergoing total hip or knee arthroplasty. Rosemont (IL): AAOS, 2007 May [online]. Available from URL: http://www.aaos.org/Research/guidelines/PE_guideline.pdf [Accessed 2007 Jul 10]
6. Bischof M, Leuppi JD, Sendi P. Cost-effectiveness of extended venous thromboembolism prophylaxis with fondaparinux in hip surgery patients. Exp Rev Pharmacoeconomics Outcomes Res 2006; 6 (2): 171-80
7. Szucs TD, Kaiser WE, Mahler F, et al. Thromboembolic prophylaxis with fondaparinux in major orthopaedic surgery: outcomes and costs. Heartdrug 2005; 5 (3): 121-30
8. Bjorvatn A, Kristiansen F. Fondaparinux sodium compared with enoxaparin sodium: a cost-effectiveness analysis. Am J Cardiovasc Drugs 2005; 5 (2): 121-30
9. Skedgel C, Goeree R, Pleasance S, et al. The cost-effectiveness of extended-duration antithrombotic prophylaxis after total hip arthroplasty. J Bone Joint Surg Am 2007; 89 (4): 819-28
10. Alton V, Eckerlund I, Norlund A. Health economic evaluations: how to find them. Int J Technol Assess Health Care 2006; 22 (4): 512-7
11. National Institute for Health Research. Centre for Reviews and Dissemination. Economic evaluation database [online]. Available from URL: http://www.crd.york.ac.uk/crdweb/ [Accessed 2008 Nov 17]

12. Russell LB, Gold MR, Siegel JE, et al. The role of cost-effectiveness analysis in health and medicine. Panel on Cost-Effectiveness in Health and Medicine. JAMA 1996 Oct 9; 276 (14): 1172-7
13. Siegel JE, Weinstein MC, Russell LB, et al. Recommendations for reporting cost-effectiveness analyses. Panel on Cost-Effectiveness in Health and Medicine. JAMA 1996 Oct 23-30; 276 (16): 1339-41
14. Weinstein MC, Siegel JE, Gold MR, et al. Recommendations of the Panel on Cost-Effectiveness in Health and Medicine. JAMA 1996 Oct 16; 276 (15): 1253-8
15. Bank of Canada. Consumer price index 1995 to present. 2008 Jun 16 [online]. Available from URL: http://www.bankofcanada.ca/en/cpi.html [Accessed 2008 Jul 12]
16. European Commission. Harmonised indices of consumer prices. 2009 [online]. Available from URL: http://epp.eurostat.ec.europa.eu/portal/page/portal/hicp/data/main_tables [Accessed 2009 Nov 22]
17. WHO. CHOosing Interventions that are Cost Effective (WHO-CHOICE) purchasing power parity 2005 [online]. Available from URL: http://www.who.int/choice/costs/ppp/en/index.html [Accessed 2009 Nov 22]
18. US Department of Labor. Bureau of Labor Statistics. Consumer price index. 2008 [online]. Available from URL: http://www.bls.gov/cpi/ [Accessed 2008 Jun 13]
19. Bachmann MO. Cost effectiveness of community-based therapeutic care for children with severe acute malnutrition in Zambia: decision tree model. Cost Eff Resour Alloc 2009; 7 (1): 2
20. Drummond MF, O'Brien B, Stoddart G, et al. Methods for economic evaluation of health care programmes. Oxford: Oxford University Press, 1997
21. Brauer CA, Rosen AB, Olchanski NV, et al. Cost-utility analyses in orthopaedic surgery. J Bone Joint Surg Am 2005 Jun; 87 (6): 1253-9
22. Hunink M, Glasziou P, Siegel J. Decision making in health and medicine. 1st ed. Cambridge: Cambridge University Press, 2001
23. Bell CM, Urbach DR, Ray JG, et al. Bias in published cost effectiveness studies: systematic review. BMJ 2006; 332 (7543): 699-703
24. Abdool-Carrim T, Adler H, Becker P, et al. The cost and benefit of prophylaxis against deep vein thrombosis in elective hip replacement. DVT/PE Prophylaxis Consensus Forum. S Afr Med J 1997; 87 (5): 594-600
25. Sarasin FP, Bounameaux H. Out of hospital antithrombotic prophylaxis after total hip replacement: low-molecular-weight heparin, warfarin, aspirin or nothing? A cost-effectiveness analysis. Thromb Haemost 2002; 87 (4): 586-92
26. Anderson DR, O'Brien B, Nagpal S. Economic evaluation comparing low molecular weight heparin with other modalities for the prevention of deep vein thrombosis and pulmonary embolism following total hip or knee arthroplasty. Ottawa (ON): Canadian Coordinating Office for Health Technology Assessment (CCOHTA), 1998
27. Bell GK, Goldhaber SZ. Cost implications of low molecular weight heparins as prophylaxis following total hip and knee replacement. Vasc Med 2001; 6 (1): 23-9
28. Botteman MF, Caprini J, Stephens JM, et al. Results of an economic model to assess the cost-effectiveness of enoxaparin, a low-molecular-weight heparin, versus warfarin for the prophylaxis of deep vein thrombosis and associated long-term complications in total hip replacement surgery in the United States. Clin Therap 2002; 24 (11): 1960-86
29. Caprini JA, Arcelus JI, Kudrna JC, et al. Cost-effectiveness of venous thromboembolism prophylaxis after total hip replacement. Phlebology 2002; 17 (3-4): 126-33
30. Dahl OE, Pleil AM. Investment in prolonged thromboprophylaxis with dalteparin improves clinical outcomes after hip replacement. J Thromb Haemost 2003; 1 (5): 896-906
31. Francis CW, Pleil AM, Reinhart SP, et al. A pharmacoeconomic evaluation of low-molecular-weight heparin in patients after total hip-replacement surgery. P T 1999; 24 (3): 136-45
32. Friedman RJ, Dunsworth GA. Cost analyses of extended prophylaxis with enoxaparin after hip arthroplasty. Clin Orthop Relat Res 2000; 370: 171-82
33. Hawkins DW, Langley PC, Krueger KP. A pharmacoeconomic assessment of enoxaparin and warfarin as prophylaxis for deep vein thrombosis in patients undergoing knee replacement surgery. Clin Therap 1998; 20 (1): 182-95
34. Hull RD, Raskob GE, Pineo GF, et al. Subcutaneous low-molecular-weight heparin vs warfarin for prophylaxis of deep vein thrombosis after hip or knee implantation: an economic perspective. Arch Int Med 1997; 157 (3): 298-303
35. Nerurkar J, Wade WE, Martin BC. Cost/death averted with venous thromboembolism prophylaxis in patients undergoing total knee replacement or knee arthroplasty. Pharmacotherapy 2002; 22 (8): 990-1000
36. Wade WE, Hawkins DW. Cost effectiveness of outpatient anticoagulant prophylaxis after total hip arthroplasty. Orthopedics 2000; 23 (4): 335-8
37. Wade WE, Spruill WJ. Cost analysis of the American College of Chest Physicians guidelines for deep vein thrombosis prophylaxis in patients undergoing orthopedic arthroplastic surgery. Pharmacotherapy 1997; 17 (6): 1286-91
38. Gordois A, Posnett J, Borris L, et al. The cost-effectiveness of fondaparinux compared with enoxaparin as prophylaxis against thromboembolism following major orthopedic surgery. J Thromb Haemost 2003; 1 (10): 2167-74
39. Dranitsaris G, Kahn SR, Stumpo C, et al. Pharmacoeconomic analysis of fondaparinux versus enoxaparin for the prevention of thromboembolic events in orthopedic surgery patients. Am J Cardiovasc Drugs 2004; 4 (5): 325-33
40. Wade WE, Spruill WJ, Leslie RB. Cost analysis: fondaparinux versus preoperative and postoperative enoxaparin as venous thromboembolic event prophylaxis in elective hip arthroplasty. Am J Orthop 2003; 32 (4): 201-5
41. Spruill WJ, Wade WE, Leslie RB. Cost analysis of fondaparinux versus enoxaparin as venous thromboembolism prophylaxis in elective hip replacement surgery. Blood Coagul Fibrinolysis 2004; 15 (7): 539-43
42. Annemans L, Minjoulat-Rey MC, De Knock M, et al. Cost consequence analysis of fondaparinux versus enoxaparin in the prevention of venous thromboembolism after major orthopaedic surgery in Belgium. Acta Clinica Belgica 2004; 59 (6): 346-57

43. Spruill WJ, Wade WE, Leslie RB. A cost analysis of fondaparinux versus enoxaparin in total knee arthroplasty. Am J Thera 2004; 11 (1): 3-8
44. Lundkvist J, Bergqvist D, Jonsson B. Cost-effectiveness of fondaparinux vs enoxaparin as venous thromboembolism prophylaxis in Sweden. Eur J Health Econ 2003; 4 (4): 254-62
45. Sullivan SD, Davidson BL, Kahn SR, et al. A cost-effectiveness analysis of fondaparinux sodium compared with enoxaparin sodium as prophylaxis against venous thromboembolism: use in patients undergoing major orthopaedic surgery. Pharmacoeconomics 2004; 22 (9): 605-20
46. Bergqvist D. Cost-effectiveness of prolonged administration of a low molecular weight heparin for the prevention of deep venous thrombosis following total hip replacement. Value Health 1999; 2 (4): 288-94
47. Davies LM, Richardson GA, Cohen TA. Economic evaluation of enoxaparin as postdischarge prophylaxis for deep vein thrombosis (DVT) in elective hip surgery. Value Health 2000; 3 (6): 397-406
48. Detournay B, Planes A, Vochelle N, et al. Cost effectiveness of a low-molecular-weight heparin in prolonged prophylaxis against deep vein thrombosis after total hip replacement. Pharmacoeconomics 1998; 13 (1 Pt 1): 81-9
49. Haentjens P, De Groote K, Annemans L. Prolonged enoxaparin therapy to prevent venous thromboembolism after primary hip or knee replacement: a cost-utility analysis. Arch Orthop Trauma 2004; 124 (8): 507-17
50. Dranitsaris G, Stumpo C, Smith R, et al. Extended dalteparin prophylaxis for venous thromboembolic events: cost-utility analysis in patients undergoing major orthopedic surgery. Am J Cardiovasc Drugs 2009; 9 (14): 45-8
51. McCullagh L, Tilson L, Walsh C, et al. A cost-effectiveness model comparing rivaroxaban and dabigatran etexilate with enoxaparin sodium as thromboprophylaxis after total hip and total knee replacement in the Irish healthcare setting. Pharmacoeconomics 2009; 27 (10): 829-46
52. Wolowacz SE, Roskell NS, Maciver F, et al. Economic evaluation of dabigatran etexilate for the prevention of venous thromboembolism after total knee and hip replacement surgery. Clin Therap 2009; 31 (1): 194-212
53. Sullivan SD, Kahn SR, Davidson BL, et al. Measuring the outcomes and pharmacoeconomic consequences of venous thromboembolism prophylaxis in major orthopaedic surgery. Pharmacoeconomics 2003; 21 (7): 477-96
54. Anderson Jr FA, Audet AM. Physician practices in the prevention of deep vein thrombosis: the MassPRO DVT Study. Orthopedics 1996; 19: 9-11
55. Anderson Jr FA, Hirsh J, White K, et al. Temporal trends in prevention of venous thromboembolism following primary total hip or knee arthroplasty 1996–2001: findings from the Hip and Knee Registry. Chest 2003 Dec; 124 (6 Suppl.): 349-56S
56. Ivanovic N, Beinema M, Brouwers JR, et al. Thromboprophylaxis in total hip-replacement surgery in Europe: acenocoumarol, fondaparinux, dabigatran and rivaroxban. Exp Rev Pharmacoeconomics Outcomes Res 2007; 7 (1): 49-58
57. Wolowacz SE, Hess N, Brennan VK, et al. Cost-effectiveness of venous thromboembolism prophylaxis in total hip and knee replacement surgery: the evolving application of health economic modelling over 20 years. Curr Med Res Opin 2008; 24 (10): 2993-3006
58. Bauer KA, Eriksson BI, Lassen MR, et al., Steering Committee of the Pentasaccharide in Major Knee Surgery Study. Fondaparinux compared with enoxaparin for the prevention of venous thromboembolism after elective major knee surgery. N Engl J Med 2001 Nov 1; 345 (18): 1305-10
59. Eriksson BI, Bauer KA, Lassen MR, et al., Steering Committee of the Pentasaccharide in Hip-Fracture Surgery Study. Fondaparinux compared with enoxaparin for the prevention of venous thromboembolism after hip fracture surgery. N Engl J Med 2001 Nov 1; 345 (18): 1298-304
60. Lassen MR, Bauer KA, Eriksson BI, et al., European Pentasaccharide Elective Surgery Study Steering Committee. Postoperative fondaparinux versus preoperative enoxaparin for prevention of venous thromboembolism in elective hip-replacement surgery: a randomised double-blind comparison. Lancet 2002 May 18; 359 (9319): 1715-20
61. Turpie AG, Bauer KA, Eriksson BI, et al. Postoperative fondaparinux versus postoperative enoxaparin for prevention of venous thromboembolism after elective hip-replacement surgery: a randomised double-blind trial. Lancet 2002 May 18; 359 (9319): 1721-6
62. Turpie AG, Bauer KA, Eriksson BI, et al. Fondaparinux vs enoxaparin for the prevention of venous thromboembolism in major orthopedic surgery: a meta-analysis of 4 randomized double-blind studies. Arch Int Med 2002; 162 (16): 1833-40
63. Quinlan DJ, Eikelboom JW, Dahl OE, et al. Association between asymptomatic deep vein thrombosis detected by venography and symptomatic venous thromboembolism in patients undergoing elective hip or knee surgery. J Thromb Haemost 2007; 5 (7): 1438-43
64. Bergqvist D, Benoni G, Bjorgell O, et al. Low-molecular-weight heparin (enoxaparin) as prophylaxis against venous thromboembolism after total hip replacement. N Engl J Med 1996 Sep 5; 335 (10): 696-700
65. Comp PC, Spiro TE, Friedman RJ, et al. Prolonged enoxaparin therapy to prevent venous thromboembolism after primary hip or knee replacement. J Bone Joint Surg Am 2001 March 1; 83 (3): 336-45
66. Dahl OE, Andreassen G, Aspelin T, et al. Prolonged thromboprophylaxis following hip replacement surgery: results of a double-blind, prospective, randomised, placebo-controlled study with dalteparin (Fragmin). Thromb Haemost 1997 Jan; 77 (1): 26-31
67. Heit JA, Elliott CG, Trowbridge AA, et al. Ardeparin sodium for extended out-of-hospital prophylaxis against venous thromboembolism after total hip or knee replacement: a randomized, double-blind, placebo-controlled trial. Ann Intern Med 2000 Jun 6; 132 (11): 853-61
68. Hull RD, Pineo GF, Francis C, et al. Low-molecular-weight heparin prophylaxis using dalteparin extended out-of-hospital vs in-hospital warfarin/out-of-hospital placebo in hip arthroplasty patients: a double-blind, randomized comparison. Arch Intern Med 2000 Jul 24; 160 (14): 2208-15
69. Planes A, Vochelle N, Darmon JY, et al. Risk of deep-venous thrombosis after hospital discharge in patients having undergone total hip replacement: double-blind

randomised comparison of enoxaparin versus placebo. Lancet 1996 Jul 27; 348 (9022): 224-8

70. Prandoni P, Bruchi O, Sabbion P, et al. Prolonged thromboprophylaxis with oral anticoagulants after total hip arthroplasty: a prospective controlled randomized study. Arch Intern Med 2002 Sep 23; 162 (17): 1966-71
71. Drummond M, McGuire A, editors. Economic evaluation in health care: merging theory with practice. New York: Oxford University Press, 2001
72. Reinhardt U. Cost-effectiveness analysis and US health care. New York Times 2009 Mar 13 [online]. Available from URL: http://economix.blogs.nytimes.com/2009/03/13/cost-effectiveness-analysis-and-us-health-care/ [Accessed 2010 Mar 16]
73. Branson BM, Handsfield HH, Lampe MA, et al. Revised recommendations for HIV testing of adults, adolescents, and pregnant women in health-care settings. MMWR Recomm Rep 2006 Sep 22; 55 (RR-14): 1-17
74. US Preventive Services Task Force. Screening for colorectal cancer: US Preventive Services Task Force recommendation statement. Ann Int Med 2008 Nov 4; 149 (9): 627-37

Correspondence: Dr *Alok Kapoor*, Hospital Medicine Unit, Boston University School of Medicine, Boston, MA 02118, USA.
E-mail: alok.kapoor@bmc.org

BMC Geriatrics

RESEARCH ARTICLE Open Access

Risk of venous thromboembolism after total hip and knee replacement in older adults with comorbidity and co-occurring comorbidities in the Nationwide Inpatient Sample (2003-2006)

Alok Kapoor[1*], Alan J Labonte[2†], Michael R Winter[2†], Jodi B Segal[3†], Rebecca A Silliman[4,5†], Jeffrey N Katz[6,7†], Elena Losina[7,8,9†], Dan Berlowitz[10,11†]

Abstract

Background: Venous thromboembolism is a common, fatal, and costly injury which complicates major surgery in older adults. The American College of Chest Physicians recommends high potency prophylaxis regimens for individuals undergoing total hip or knee replacement (THR or TKR), but surgeons are reluctant to prescribe them due to fear of excess bleeding. Identifying a high risk cohort such as older adults with comorbidities and co-occurring comorbidities who might benefit most from high potency prophylaxis would improve how we currently perform preoperative assessment.

Methods: Using the Nationwide Inpatient Sample, we identified older adults who underwent THR or TKR in the U.S. between 2003 and 2006. Our outcome was VTE, including any pulmonary embolus or deep venous thrombosis. We performed multivariate logistic regression analyses to assess the effects of comorbidities on VTE occurrence. Comorbidities under consideration included coronary artery disease, congestive heart failure (CHF), chronic obstructive pulmonary disease (COPD), diabetes, and cerebrovascular disease. We also examined the impact of co-occurring comorbidities on VTE rates.

Results: CHF increased odds of VTE in both the THR cohort (OR = 3.08 95% CI 2.05-4.65) and TKR cohort (OR = 2.47 95% CI 1.95-3.14). COPD led to a 50% increase in odds in the TKR cohort (OR = 1.49 95% CI 1.31-1.70). The data did not support synergistic effect of co-occurring comorbidities with respect to VTE occurrence.

Conclusions: Older adults with CHF undergoing THR or TKR and with COPD undergoing TKR are at increased risk of VTE. If confirmed in other datasets, these older adults may benefit from higher potency prophylaxis.

Background

Medical injury, or harm associated with a therapeutic or diagnostic intervention [1], commonly complicates major surgery [2]. Age alone confers a small effect on postoperative injury but medical comorbidities and co-occurring comorbidities contribute substantially. Co-occurring comorbidities, that is, 2 comorbidities which frequently occur together, are part of a larger phenomenon of multiple morbidity [3].

Venous thromboembolism (VTE) is a common, often fatal, and costly injury which complicates major surgery in older adults. VTE is particularly common following total hip and knee replacement (THR and TKR), with venographic rates of up to 60% without prophylaxis [4]. Newer, more potent prophylaxis regimens including the synthetic pentasacchride fondaparinux and twice daily dosing of enoxaparin, a low molecular weight heparin, offer the ability to significantly reduce the risk of VTE. Their wide spread adoption has been slow, however, given the increased risk of bleeding [5]. The ability to identify a high risk cohort among older adults undergoing THR and TKR, who would potentially benefit

* Correspondence: alok.kapoor@bmc.org
† Contributed equally
[1]Hospital Medicine Unit, Boston University School of Medicine, 715 Albany Street, Boston, 02118, MA, USA
Full list of author information is available at the end of the article

from high potency prophylaxis, would be an improvement in the way clinicians currently conduct preoperative assessments.

Comorbidities such as congestive heart failure (CHF) and chronic obstructive pulmonary disease (COPD) have been associated with increased VTE risk in some studies [6-8] but not in others [9,10]. The association with CHF may relate to blood flow stasis which is part of Virchow's triad [11] of alteration in blood flow (stasis), endothelial injury, and alterations in blood constituents (hypercoagulable factors). Similarly the association of COPD may be explained through an immobility or stasis mechanism. Newer evidence, in both surgical [6,7,9,10] and non-surgical settings [12], suggests that atherosclerotic conditions such as coronary artery disease (CAD) and cerebrovascular disease (CVD) are also associated with an increased risk of VTE. In general, these studies had a small number of observations and/or did not focus on discrete surgical procedures, making it difficult to draw conclusions about modest but clinically important effects on VTE of comorbidities for individual surgeries. The combined effect of comorbidities, i.e., co-occurring comorbidities, has been incompletely evaluated.

In this investigation, we assessed the Nationwide Inpatient Sample to determine the risk of VTE in older adults with prevalent comorbidities and co-occurring comorbidities undergoing THR or TKR. We hypothesized that CHF, CAD, CVD, and COPD would predict increased risk of VTE. We also proposed that the joint effect of co-occurring comorbidities – composed of comorbidities with distinct pathogenesis, such as CAD and COPD – would equal the sum of the component effects or possibly exceed their sum, indicating a positive interaction.

Methods

Data Sources and Study Sample

We used the Nationwide Inpatient Sample (NIS) from 2003-2006 for our analysis [13]. This work involves the use of publicly available, archival information abstracted from patient medical records and was approved by the Institutional Review Board at the Boston University Medical Center in Boston, Massachusetts. The NIS contains information from hospital inpatient stays collected at the state-level and then assembled by the Agency for Healthcare Research and Quality (AHRQ) as part of the Healthcare Utilization Project (HCUP). NIS includes primary and secondary diagnoses, procedures, admission and discharge status, patient demographics (gender, age, race, median income, and residence zip code), expected payment source, total charges, length of stay, and hospital characteristics (ownership, size, and teaching status). We identified a cohort of patients (age 65 or older) who underwent primary THR or TKR identified through ICD-9-CM procedure codes 81.51 and 81.54. We restricted our analysis to elective, primary THR and TKR because we felt that non-elective surgery and revision joint surgery represented a distinct population with different risks for VTE. We therefore excluded surgical cases conducted in the setting of fracture, pelvic or thigh infection, or removal of prior prosthesis, internal fixation device, implant or graft consistent with earlier work [14,15] conducted by members of our team.

Outcomes and Exposures

We determined VTE to have occurred if any one of eight ICD-9-CM DVT codes, three PE codes, and two non-specific VTE codes were present at discharge (Additional file 1).

We included specific comorbidities based on evidence of associations with VTE documented in the literature or a known biological link to the outcomes These included CAD, CHF, COPD, CVD, and diabetes. Comorbidities and presence of obesity were coded as by Elixhauser et al.[16] (Additional file 1). Multiple publications [17-19] have supported the use of the Elixhauser comorbidity coding algorithms rather than the earlier Charlson comorbidity index, including those looking at discrete surgical procedures [20,21]. The accuracy of coding of the comorbidities has been validated in multiple publications [22,23]. We also examined co-occurring comorbidities from the above mentioned comorbidities if they had a prevalence of greater than 2% (given that exposures occurring less frequently than 2% of the time would be infeasible to study because of sample size constraints).

Consistent with previous work [14,15,24], we evaluated the effects of potentially confounding factors, i.e., race, insurance status, hospital surgical volume, obesity, bi-laterality (two primary procedures during the same surgery), chronic kidney disease, and hypercoagulable state (including cancer and genetic predisposition), by adding them, one at a time, to the crude model (with comorbidity exposure, age group, and sex). If the addition of a given factor did not change the odds ratios for the comorbidity exposure by more than 10%, it was deleted from the final model.

Analysis

We calculated descriptive statistics and cross-tabular frequencies for each outcome and comorbidity to determine the shape of distributions, the extent of missing data, and the presence of small frequencies of the discrete comorbidity levels. We examined comorbidities for statistical interactions with each other. Specifically, we measured whether prevalent combinations of comorbidities (as defined above) produced statistically significant associations with VTE in multivariate models. Because we found

these to be present in the TKR cohort, for the ease of interpretability, we report the associations between co-occurring comorbidities and outcomes using a categorical comorbidity variable for both the THR and TKR cohorts. We created a ten level categorical variable with a separate, mutually exclusive, level for each of nine comorbidities or co-occurring comorbidities (with 1 additional for all other combinations of two or more comorbidities). While this limited the population size for each comorbidity group, it allowed us to compare the risk of VTE for groups of older adults with single or co-occurring comorbidities against a common reference group of older adults without any of the nine comorbidities.

Finally, we built multivariate logistic regression models to assess the independent effects of comorbidities. In our analysis we retained age and sex in all models. For remaining potential confounders, we retained any variable that changed the effect estimate for any comorbidity by more than 10%.

Results

In NIS 2003-2006, we identified 93,071 primary THR and 223,600 primary TKR surgeries. Overall, the age of older adults was evenly distributed into quartiles 65-69, 70-74, 75-79, and 80 or older. Sixty-three percent of subjects were female. (Table 1) CAD, COPD, and diabetes alone occurred most frequently (9.8%, 8.6%, and 8.4%, respectively for the THR cohort). CVD and CHF occurred much less frequently (0.6% and 1.2%, respectively for the THR cohort). (Table 1) CAD with diabetes was present in 2.5%, and CAD with COPD was present in 1.9% of the THR cohort. (Table 1) Other combinations occurred approximately 1% of the time. Similar trends occurred in the TKR cohort.

VTE during the index hospitalization occurred 0.8% of the time after THR and 1.2% of the time after TKR. In the THR cohort, the rate of VTE ranged from 0.6% in older adults with diabetes alone to 2.3% in older adults with CHF alone; the rate of VTE in older adults without any of the identified comorbidities was 0.7%. (Table 2) Similarly, in the TKR cohort, this range was 1.1% to 2.6%; the rate of VTE in older adults without any of the identified comorbidities was 1.1%. In multivariate analysis CHF predicted a threefold increase in the odds of VTE in the THR cohort (OR = 3.08 95% CI 2.05-4.65) and a similar increase in the TKR cohort (OR = 2.47 95% CI 1.95-3.14). (Table 3) COPD predicted a 50% increase in odds in the TKR cohort (OR = 1.49 95% CI 1.31-1.70). (Table 3)

We did not find any positive interactions between comorbidities in our analysis. The combination of CAD and CHF alone was associated with a twofold increase in odds in both the THR and TKR cohort, an effect which was smaller than for CHF alone. The combination of diabetes and CAD was associated with a 39% decrease in the odds of VTE only in the TKR cohort, an effect which was smaller, i.e., more protective, than either comorbidity alone. (Table 3) The six potential confounding factors that we tested were deleted from the final model for lack of confounding.

Discussion

We comprehensively examined the association of comorbidities and co-occurring comorbidities and VTE in a unique population of older adults undergoing primary total hip and knee replacement, high risk surgeries for VTE. We found that the rates of VTE captured in administrative data of older adults for the period immediately following THR and TKR to be low at 0.8-1.2%. Having CHF substantially increased the odds of VTE after THR or TKR and having COPD somewhat increased the odds of VTE after TKR. Co-occurring comorbidities did not increase the risk of VTE beyond their individual effects.

Comparison of our results with previous studies is limited by differences in outcomes measured and populations studied. Gangireddy et al. [9] conducted one of the largest studies to date using data from the Veterans Affairs National Surgical Quality Improvement Program (NSQIP), which included veterans undergoing nine different surgeries, including THR, between 1996 and 2001. After controlling for multiple preoperative and postoperative clinical variables, a multivariate analysis with 76,771 individuals showed that CHF and COPD were not associated with increased rates of VTE. This study supported our findings of the association between diabetes and a slightly lower rate of VTE (OR = 0.75).

We also detected a 29% reduction in the risk of VTE in patients undergoing THR with CVD. This result was not statistically significant nor did we detect an association in the knee population. There has been little evidence, however, regarding the relationship between CVD and postoperative VTE. Prior work [12] has suggested a common inflammatory pathway but this has not been evaluated extensively in the postoperative setting. In our study the association was not present for both knee and hip cohorts and was not statistically significant. We plan to re-assess this relationship in our future work.

Kikura et al. [6] examined 21,903 Japanese patients of multiple ages and multiple surgery types and found that history of acute myocardial infarction (AMI) was significantly related (OR = 7.7 95% CI 1.7-34.7) to the development of postoperative thrombotic events (including repeat AMI). Although we did collect information about the history of AMI in particular, we did not find an association with CAD in general in our analysis. In a cohort of 269 post-menopausal women undergoing THR and TKR, Jaffer et al. [7] found a trend towards

Table 1 Frequency counts (and row percentages) for selected variables and stratified by surgical procedure

	Total Hip Replacement	Total Knee Replacement
Age		
65-69 yr	23405 (25.2)	63363 (28.3)
70-74 yr	24888 (26.7)	63608 (28.5)
75-79 yr	23316 (25.0)	54578 (24.4)
≥80 yr	21462 (23.1)	42051 (18.8)
Gender		
Male	34578 (37.2)	78539 (35.1)
Female	58492 (62.8)	145061 (64.9)
Comorbid Diseases		
Cerebrovascular Disease (CVD) alone	570 (0.6)	1356 (0.6)
Coronary Artery Disease (CAD) alone	9132 (9.8)	20017 (9.0)
Congestive Heart Failure (CHF) alone	1097 (1.2)	2746 (1.2)
Diabetes alone	7782 (8.4)	26740 (12.0)
Chronic Obstructive Pulmonary Disease (COPD) alone	7989 (8.6)	16836 (7.5)
CAD and CHF alone	731 (0.8)	1678 (0.8)
CAD and COPD alone	1731 (1.9)	3366 (1.5)
CAD and Diabetes alone	2320 (2.5)	7155 (3.2)
COPD and Diabetes alone	1181 (1.3)	3688 (1.7)
All other 2+ combinations	2725 (2.9)	6878 (3.1)
None of the above comorbidities	57813 (62.0)	133140 (59.4)
Median Length of Stay (days)	4	3

Table 2 Frequency counts for selected variables stratified by age and surgical procedure

	Age Group									
	Hip Procedures					Knee Procedures				
Exposure	**65-69 years of age**	**70-74 years of age**	**75-79 years of age**	**≥80 years of age**	**total**	**65-69 years of age**	**70-74 years of age**	**75-79 years of age**	**≥80 years of age**	**total**
CVD alone	87	129	160	194	570	241	342	382	391	1356
CAD alone	1677	2269	2503	2683	9132	4065	5413	5608	4931	20017
CHF alone	143	209	277	468	1097	441	574	702	1029	2746
Diabetes alone	2159	2229	1935	1459	7782	8975	8197	6116	3452	26740
COPD alone	2110	2220	1949	1710	7989	4978	4873	4065	2920	16836
CAD and CHF alone	87	124	190	330	731	249	335	449	645	1678
CAD and COPD alone	359	443	485	444	1731	726	937	978	725	3366
CAD and Diabetes alone	512	673	632	503	2320	1909	2216	1906	1124	7155
COPD and Diabetes alone	341	356	298	186	1181	1359	1148	772	409	3688
All other 2+ combinations	482	673	763	807	2725	1593	1904	1896	1485	6878
None of the above comorbidities	15448	15563	14124	12678	57813	38827	37669	31704	24940	133140

Abbreviations: CAD = coronary artery disease, CHF = congestive heart failure, COPD = chronic obstructive pulmonary disease, CVD = cerebrovascular disease.

Table 3 Odds ratios (and 95% CIs) for exposures, adjusted† for covariates and stratified by surgical procedure

	HIP PROCEDURES				KNEE PROCEDURES			
Exposure	**VTE yes**	**VTE no**	**OR†**	**CIs**	**VTE yes**	**VTE no**	**OR**	**CIs**
No Comorbidity	410	57403	**1.00**	**(ref)**	1458	131682	**1.00**	**(ref)**
CVD vs.	3	567	**0.71**	(0.23-2.23)	21	1335	**1.42**	(0.92-2.20)
CAD vs.	68	9064	**0.98**	(0.76-1.28)	257	19760	**1.19**	(1.04-1.37)
CHF vs.	25	1072	**3.08**	(2.05-4.65)	73	2673	**2.47**	(1.95-3.14)
Diabetes vs.	47	7735	**0.85**	(0.63-1.15)	282	26458	**0.96**	(0.85-1.10)
COPD vs.	60	7929	**1.06**	(0.81-1.39)	273	16563	**1.49**	(1.31-1.70)
CAD and CHF vs.	12	719	**2.11**	(1.18-3.78)	33	1645	**1.86**	(1.30-2.60)
CAD & COPD vs.	16	1715	**1.25**	(0.75-2.06)	52	3314	**1.44**	(1.09-1.90)
Diabetes & CAD vs.	18	2302	**1.05**	(0.65-1.69)	47	7108	**0.61**	(0.45-0.81)
Diabetes & COPD vs.	9	1172	**1.09**	(0.56-2.11)	36	3652	**0.89**	(0.64-1.24)
All other combinations	34	2691	**1.69**	(1.19-2.40)	111	6767	**1.49**	(1.23-181)
Age 65-69	163	23242	**1.00**	**(ref)**	734	62629	**1.00**	**(ref)**
Age 70-74 vs. Age 65-69	167	24721	**0.96**	(0.77-1.19)	751	62857	**1.01**	(0.91-1.12)
Age 75-79 vs. Age 65-69	172	23144	**1.05**	(0.84-1.30)	649	53929	**1.00**	(0.90-1.12)
Age ≥80 vs. Age 65-69	200	21262	**1.30**	(1.05-1.60)	509	41542	**0.99**	(0-89-1.11)
Females	411	58081	**1.00**	**(ref)**	1746	143315	**1.00**	**(ref)**
Males	291	34287	**1.21**	(1.04-1.41)	897	77642	**0.94**	(0.87-1.02)
Total observations	702	92369			2643	220957		

†Associations adjusted in multivariate logistic regression for each comorbidity exposure as compared with the group with none of the listed comorbidities or co-occurring comorbidities. All association were adjusted for age group and sex. We also assessed race, insurance status, hospital surgical volume, obesity, bi-laterality (two primary procedures during the same surgery) and hypercoagulable state but deleted them from the final model for lack of confounding.
Abbreviations: CAD = coronary artery disease, CHF = congestive heart failure, COPD = chronic obstructive pulmonary disease, CVD = cerebrovascular disease.

CHF predicting more postoperative VTE events (OR = 5.50 95% CI 0.94-43.3) but a trend towards COPD predicting fewer VTE events (OR = 0.42 95% CI 0.07-1.98).

The association of CHF and VTE may relate to blood flow stasis as discussed earlier. Alternatively, CHF may indicate a degree of immobility that was not measured in the data we analyzed. Other comorbidities may also contribute to the development of postoperative VTE but their effects may have been attenuated by a selection bias. Surgeons may select only the healthiest subset of older adults with comorbidities for surgery. The absence of positive interactions between frequently co-occurring comorbidities (especially CAD and CHF) also suggests a potential source of a selection bias. Older adults with co-occurring comorbidities deemed to be suitable surgical candidates are presumably healthier in other ways than other older adults with the same comorbidities.

In the case of COPD, we only detected an increase in risk for older adults undergoing knee surgery. This could be explained by the generally weak predictor effect of COPD on VTE or it could be related to the inherent differences between hip and knee surgery. Postoperative mobility may be significantly less for hip surgery and the effect of immobility in this group may dwarf other predictors such as COPD. Future work should examine the interaction between mobility and surgery type in data where this information is available.

The association of CHF and COPD with postoperative VTE has important implications. Although the American College of Chest Physicians currently recommends high potency prophylaxis such as fondaparinux or LMWH for all individuals undergoing THR and TKR [4], surgeons are reluctant to prescribe these regimens, fearing bleeding complications. Identification of a high risk subset among a group of older adults already at increased risk for VTE based on the surgery planned would be an important improvement in the way we currently perform preoperative assessment. In addition to the use of high potency prophylaxis, surgeons may also use the risk information to incorporate other practices, such as regional anesthesia, mechanical prophylaxis devices, or stockings, aimed at lowering VTE rates.

There are limitations to the work we presented. Due to the nature of the NIS administrative data we have limited ability to capture VTE. A recent study [24], suggests that administrative data capture only 58% of VTE events. There is no evidence, however, to suggest that the events indentified are differentially being diagnosed in individuals with CHF or other comorbidities. In addition, we did not have access to medication information including prophylaxis agent. A recent study by Cohen et al. in 2008 [25] indicated that in the United States, only 48% of medical patients are receiving the recommended ACCP prophylaxis and only 71% of surgical patients are

receiving prophylaxis [25]. If comorbidities prompted physicians to prescribe more potent prophylaxis in older adults with CHF or other comorbidities, however, the effects we observed would represent an underestimate of the true effect. A recent survey [26] suggests that orthopedic surgeons vary their prescribing patterns less than 10% of the time when evaluating a patient with cardiopulmonary disease. Future work should examine the relationship between comorbidities and VTE while controlling for prophylaxis agents in data where medication information is available.

We did not have information about events which took place after hospitalization. Given that the median time for development of DVT is 17 days for THR and 7 days for TKR [27] and the median length of stay was 3 or 4 days for each surgery in our analysis, the associations we present may not reflect the experience of older adults who develop injury in the post discharge period. Controlling for length of stay would not disentangle the relation between these comorbidities and VTE and we, therefore, did not control for it in our analysis. Length of stay may very well be a surrogate for immobility and stasis which are on the causal pathway of VTE development. Alternatively, increased length of stay may also be associated with VTE because of added time needed to achieve therapeutic levels of warfarin. Even though NIS data does not allow for measurement of the 30 or 90 day incidence of VTE, we believe that post discharge rates of VTE events will be similarly disparate in individuals with compared to those without comorbidities. In the future we plan to confirm these associations in data where this information is available.

Administrative data are susceptible to upcoding where medical coders assign a diagnosis that may have only been considered but not proven. We did not have information about those VTE events that were present on admission compared to those that occurred during hospitalization. We plan to conduct further validation studies in other databases where pre-existing diagnosis modifiers are available. We cannot firmly establish causality between comorbidities and VTE using the data available to us. Comorbidities may be linked to other processes such as increased operative time or difficulty weaning from a ventilator after surgery. Future research with datasets containing these clinical variables may clarify the exact causal pathway. Lastly, although in our analysis we controlled for the presence of several factors that might increase the risk of VTE, we did not have data on the smoking status of individuals, which might be related to both the exposure and outcome.

Conclusion

CHF is strongly associated with an increased risk of VTE after THR and TKR; COPD is associated with an increased risk of VTE after TKR. The absence of risk associated with other comorbidities and co-occurring comorbidities may be explained by the selection of healthier older adults for these surgeries or by some of the limiting factors described above. If these findings are confirmed in other datasets, higher risk older adults with comorbidities could potentially benefit from more aggressive preventive interventions including high potency pharmacologic prophylaxis.

Additional material

Additional file 1: Appendix I - ICD-9-CM Procedure and Diagnosis Codes. total Joint Replacement Procedure Codes and Venous Thromboembolism Diagnosis Codes.

Acknowledgements

This work was supported by a generous pilot grant from the Boston University School of Medicine Department of Medicine Pilot Grant Program. This work involves the use of publicly available, archival information abstracted from patient medical records and was approved by the Institutional Review Board at the Boston University Medical Center in Boston, Massachusetts.

Author details

[1]Hospital Medicine Unit, Boston University School of Medicine, 715 Albany Street, Boston, 02118, MA, USA. [2]Data Coordinating Center, Boston University School of Public Health, 715 Albany Street, Talbot Building, Boston, 02118, MA, USA. [3]Department of Medicine, Johns Hopkins University School of Medicine, 733 N. Broadway, Baltimore, 21205, MD, USA. [4]Section of Geriatrics, Boston University School of Medicine, 715 Albany Street, Boston, 02118 MA, USA. [5]Section of Epidemiology, Boston University School of Public Health, 715 Albany Street, Talbot Building, Boston, 02118, MA, USA. [6]Department of Orthopaedic Surgery and Division of Rheumatology, Immunology, and Allergy, Orthopaedic and Arthritis Center for Outcomes Research, Brigham and Women's Hospital, 75 Francis Street, Boston, 02115, MA, USA. [7]Harvard Medical School, 25 Shattuck Street, Boston, 02115, MA, USA. [8]Department of Orthopaedic Surgery, Brigham and Women's Hospital, 75 Francis Street, Boston, 02115, MA, USA. [9]Department of Biostatistics, Boston University School of Public Health, 715 Albany Street, Talbot Building, Boston, 02118, MA, USA. [10]Center for Health Quality, Outcomes, and Economic Research, Edith Nourse Rogers Memorial VA Hospital, 200 Springs Road, Bedford, 01730, MA, USA. [11]Boston University School of Public Health 715 Albany Street, Talbot Building, Boston, 02118, MA, USA.

Authors' contributions

AK and AL were responsible for the study design, analysis, interpretation, and manuscript write-up. RAS, JNK, EL, and DB were responsible for the study design, interpretation, and manuscript write-up. MW was responsible for the study analysis and interpretation, and JBS was responsible for the study interpretation and manuscript write-up. All authors read and approved the final manuscript.

Competing interests

None of the authors have any financial or non-financial competing interests to declare as related to the contents of this manuscript.

Received: 10 April 2010 Accepted: 17 September 2010
Published: 17 September 2010

References

1. Patient Safety and Quality: Not adjusting for pre-existing health problems may have exaggerated the number of deaths due to medical injury. [http://www.ahrq.gov/research/feb07/0207RA5.htm].

2. Zhan C, Miller MR: **Excess length of stay, charges, and mortality attributable to medical injuries during hospitalization.** *JAMA* 2003, **290(14)**:1868-1874.
3. Boyd CM, Weiss CO, Halter J, Han KC, Ershler WB, Fried LP: **Framework for evaluating disease severity measures in older adults with comorbidity.** *Journals of Gerontology Series A Biological Sciences & Medical Sciences* 2007, **62(3)**:286-295.
4. Geerts WH, Pineo GF, Heit JA, Bergqvist D, Lassen MR, Colwell CW, Ray JG: **Prevention of venous thromboembolism: the Eighth ACCP Conference on Antithrombotic and Thrombolytic Therapy.** *Chest* 2008, **133(3 Suppl)**:381S-453.
5. **American Academy of Orthopaedic Surgeons Clinical Guideline on Prevention of Symptomatic Pulmonary Embolism in Patients Undergoing Total Hip or Knee Arthroplasty.** [http://www.aaos.org/Research/guidelines/PE_guideline.pdf].
6. Kikura M, Takada T, Sato S: **Preexisting morbidity as an independent risk factor for perioperative acute thromboembolism syndrome.** *Arch Surg* 2005, **140(12)**:1210-1217, discussion 1218.
7. Jaffer AK, Barsoum WK, Krebs V, Hurbanek JG, Morra N, Brotman DJ: **Duration of anesthesia and venous thromboembolism after hip and knee arthroplasty.** *Mayo Clin Proc* 2005, **80(6)**:732-738.
8. Fimognari FL, Scarlata S, Conte ME, Incalzi RA, Fimognari FL, Scarlata S, Conte ME, Incalzi RA: **Mechanisms of atherothrombosis in chronic obstructive pulmonary disease.** *Int J Chron Obstruct Pulmon Dis* 2008, **3(1)**:89-96.
9. Gangireddy C, Rectenwald JR, Upchurch GR, Wakefield TW, Khuri S, Henderson WG, Henke PK: **Risk factors and clinical impact of postoperative symptomatic venous thromboembolism.** *J Vasc Surg* 2007, **45(2)**:335-341, discussion 341-332.
10. Schiff RL, Kahn SR, Shrier I, Strulovitch C, Hammouda W, Cohen E, Zukor D: **Identifying orthopedic patients at high risk for venous thromboembolism despite thromboprophylaxis.** *Chest* 2005, **128(5)**:3364-3371.
11. **Overview of the causes of venous thrombosis.** [http://www.uptodate.com].
12. Prandoni P, Bilora F, Marchiori A, Bernardi E, Petrobelli F, Lensing AW, Prins MH, Girolami A, Prandoni P, Bilora F, *et al*: **An association between atherosclerosis and venous thrombosis.** *New England Journal of Medicine* 2003, **348(15)**:1435-1441.
13. **Overview of the Nationwide Inpatient Sample (NIS).** [http://www.hcup-us.ahrq.gov/nisoverview.jsp].
14. Katz JN, Barrett J, Mahomed NN, Baron JA, Wright RJ, Losina E: **Association between hospital and surgeon procedure volume and the outcomes of total knee replacement.** *Journal of Bone & Joint Surgery - American Volume* 2004, **86-A(9)**:1909-1916.
15. Katz JN, Losina E, Barrett J, Phillips CB, Mahomed NN, Lew RA, Guadagnoli E, Harris WH, Poss R, Baron JA: **Association between hospital and surgeon procedure volume and outcomes of total hip replacement in the United States medicare population.** *Journal of Bone & Joint Surgery American* 2001, **83(11)**:1622-1629.
16. Elixhauser A, Steiner C, Harris DR, Coffey RM: **Comorbidity measures for use with administrative data.** *Med Care* 1998, **36(1)**:8-27.
17. Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi J-C, Saunders LD, Beck CA, Feasby TE, Ghali WA: **Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data.** *Med Care* 2005, **43(11)**:1130-1139.
18. Dominick KL, Dudley TK, Coffman CJ, Bosworth HB: **Comparison of three comorbidity measures for predicting health service use in patients with osteoarthritis.** *Arthritis & Rheumatism* 2005, **53(5)**:666-672.
19. Southern DA, Quan H, Ghali WA: **Comparison of the Elixhauser and Charlson/Deyo methods of comorbidity measurement in administrative data.** *Med Care* 2004, **42(4)**:355-360.
20. Livingston EH: **Development of bariatric surgery-specific risk assessment tool.** *Surg* 2007, **3(1)**:14-20, discussion 20.
21. Fry DE, Pine M, Jordan HS, Elixhauser A, Hoaglin DC, Jones B, Warner D, Meimban R: **Combining administrative and clinical data to stratify surgical risk.** *Ann Surg* 2007, **246(5)**:875-885.
22. Birman-Deych E, Waterman AD, Yan Y, Nilasena DS, Radford MJ, Gage BF: **Accuracy of ICD-9-CM Codes for Identifying Cardiovascular and Stroke Risk Factors.** *Med Care* 2005, **43(5)**:480-485.
23. Lee DS, Donovan L, Austin PC, Gong Y, Liu PP, Rouleau JL, Tu JV, Lee DS, Donovan L, Austin PC, *et al*: **Comparison of coding of heart failure and comorbidities in administrative and clinical data for use in outcomes research.** *Med Care* 2005, **43(2)**:182-188.
24. Romano PS, Mull HJ, Rivard PE, Zhao S, Henderson WG, Loveland S, Tsilimingras D, Christiansen CL, Rosen AK, Romano PS, *et al*: **Validity of selected AHRQ patient safety indicators based on VA National Surgical Quality Improvement Program data.** *Health Serv Res* 2009, **44(1)**:182-204.
25. Cohen AT, Tapson VF, Bergmann J-F, Goldhaber SZ, Kakkar AK, Deslandes B, Huang W, Zayaruzny M, Emery L, Anderson FA: **Venous thromboembolism risk and prophylaxis in the acute hospital care setting (ENDORSE study): a multinational cross-sectional study.** *Lancet* 2008, **371(9610)**:387-394.
26. Markel DC, York S, Liston MJ, Flynn JC, Barnes CL, Davis CM: **Venous Thromboembolism: Management by American Association of Hip and Knee Surgeons.** *The Journal of Arthroplasty* 2009, **25(1)**:3-9.e2.
27. White RH, Romano PS, Zhou H, Rodrigo J, Bargar W: **Incidence and Time Course of Thromboembolic Outcomes Following Total Hip or Knee Arthroplasty.** *Arch Intern Med* 1998, **158(14)**:1525-1531.

Pre-publication history

The pre-publication history for this paper can be accessed here:
http://www.biomedcentral.com/1471-2318/10/63/prepub

doi:10.1186/1471-2318-10-63
Cite this article as: Kapoor *et al.*: Risk of venous thromboembolism after total hip and knee replacement in older adults with comorbidity and co-occurring comorbidities in the Nationwide Inpatient Sample (2003-2006). *BMC Geriatrics* 2010 **10**:63.

Free Testosterone Levels Are Associated with Mobility Limitation and Physical Performance in Community-Dwelling Men: The Framingham Offspring Study

Joanne B. Krasnoff,* Shehzad Basaria,* Michael J. Pencina, Guneet K. Jasuja, Ramachandran S. Vasan, Jagadish Ulloor, Anqi Zhang, Andrea Coviello, Margaret Kelly-Hayes, Ralph B. D'Agostino, Philip A. Wolf, Shalender Bhasin,† and Joanne M. Murabito†

Boston University School of Medicine, Sections of Endocrinology, Diabetes, and Nutrition (J.B.K., S.Ba., J.U., A.Z., A.C., S.Bh.) Preventative Medicine and Epidemiology (R.S.V.) and Department of Neurology (M.K.-H., P.A.W.) and Department of Biostatistics (M.J.P.), Boston University, Boston, Massachusetts 02118; Department of Mathematics, Statistics and Consulting Unit (M.J.P., G.K.J., R.B.D.), Boston University, Boston, Massachusetts 02215; and The National Heart, Lung, and Blood Institute's Framingham Heart Study (M.J.P., G.K.J., R.S.V., M.K.-H., R.B.D., P.A.W., J.M.M.) and Boston University School of Medicine, Department of Medicine (J.M.M.), Section of General Internal Medicine, Framingham, Massachusetts 01701

Context: Mobility limitation is associated with increased morbidity and mortality. The relationship between circulating testosterone and mobility limitation and physical performance is incompletely understood.

Objective: Our objective was to examine cross-sectional and prospective relations between baseline sex hormones and mobility limitations and physical performance in community-dwelling older men.

Design, Setting, and Participants: We conducted cross-sectional and longitudinal analyses of 1445 men (mean age 61.0 ± 9.5 yr) who attended Framingham Offspring Study examinations 7 and 8 (mean 6.6 yr apart). Total testosterone (TT) was measured by liquid chromatography tandem mass spectrometry at examination 7. Cross-sectional and longitudinal analyses of mobility limitation and physical performance were performed with continuous (per SD) and dichotomized [low TT and free testosterone (FT) and high SHBG *vs.* normal] hormone levels.

Main Outcome Measures: Self-reported mobility limitation, subjective health, usual walking speed, and grip strength were assessed at examinations 7 and 8. Short physical performance battery was performed at examination 7.

Results: Higher continuous FT was positively associated with short physical performance battery score ($\beta = 0.13$; $P = 0.008$), usual walking speed ($\beta = 0.02$; $P = 0.048$), and lower risk of poor subjective health [odds ratio (OR) = 0.72; $P = 0.01$]. In prospective analysis, 1 SD increase in baseline FT was associated with lower risk of developing mobility limitation (OR = 0.78; 95% confidence interval = 0.62–0.97) and progression of mobility limitation (OR = 0.75; 95% confidence interval = 0.60–0.93). Men with low baseline FT had 57% higher odds of reporting incident mobility limitation ($P = 0.03$) and 68% higher odds of worsening of mobility limitation ($P = 0.007$).

Conclusions: Lower levels of baseline FT are associated with a greater risk of incident or worsening mobility limitation in community-dwelling older men. Whether this risk can be reduced with testosterone therapy needs to be determined by randomized trials. **(*J Clin Endocrinol Metab* 95: 2790–2799, 2010)**

ISSN Print 0021-972X ISSN Online 1945-7197
Printed in U.S.A.

doi: 10.1210/jc.2009-2680 Received December 15, 2009. Accepted March 17, 2010.
First Published Online April 9, 2010
* J.B.K. and S.Ba. are joint first authors.
† S.Bh. and J.M.M. are joint senior authors.

For editorial see page 2634

Abbreviations: BMI, Body mass index; CI, confidence interval; FHS, Framingham Heart Study; FT, free testosterone; LC-MS/MS, liquid chromatography tandem mass spectrometry; OR, odds ratio; SPPB, short physical performance battery; TT, total testosterone.

The powerful demographic shift toward aging of human populations has focused attention on remediable factors that limit the ability of older individuals to live independently. Among individuals age 65 yr and older, 44% report some mobility limitation (1). Decline in mobility is associated with loss of independence (2) and increased risk of disability (3), institutionalization (4), decreased quality of life (5), and death (3, 6). Mobility limitation in older individuals is undoubtedly multifactorial, but age-related declines in muscle mass, strength, and power are important contributors (7, 8).

Total testosterone (TT) levels in men decline progressively with age (9–14). Because SHBG increases with age (9, 10, 15), the decline in free testosterone (FT) with aging is even greater than the decline in TT levels. Age-related decline in testosterone levels has been associated with reduced muscle mass and lower extremity strength in older men (16, 17).

Only a few studies have addressed the relationships between circulating testosterone levels and mobility and physical function, and the data are conflicting. The Longitudinal Ageing Study Amsterdam reported a significant positive association between low TT and low grip strength in older men, but no statistically significant relationship with self-reported functional limitations was observed in that study (18). The Massachusetts Male Ageing Study found no relationship between TT and grip strength (19). An analysis of longitudinal data from two independent cohorts of older men showed no association between TT and FT and the decline in physical performance (20). The BACH/Bone Study concluded age-associated alterations in sex hormone levels play a minor role in age-related declines in muscle strength and physical performance (21). These studies measured testosterone levels by RIAs, whose accuracy has been questioned (22, 23). Thus, the relationships among circulating levels of sex hormones and mobility limitations and physical performance in older men are inadequately understood.

Using data from the Framingham Offspring Study, we determined whether TT, FT, and SHBG are related cross-sectionally to mobility limitations, subjective health, and performance-based measures of physical function in community-dwelling older men. Additionally, in longitudinal analyses, we evaluated whether these hormones prospectively are associated with incident mobility limitation and worsening of mobility and subjective health in older men. We measured circulating TT levels by liquid chromatography tandem mass spectrometry (LC-MS/MS), widely considered the reference method for testosterone measurement (23).

Materials and Methods

The Boston University Institutional Review Board approved the study, and written informed consent was obtained from all participants.

Study sample

The Framingham Heart Study (FHS) began in 1948 as a prospective study to examine the risk factors for heart disease (24). In 1971, enrollment of offspring of the original cohort and spouses of the offspring constituted the Framingham Offspring Study (25). The offspring cohort has completed eight examinations at approximately 4- to 8-yr intervals. The men of the Offspring cohort who attended examination 7 (1998–2001) were eligible for the present study (n = 1625). Men with prostate cancer undergoing androgen deprivation therapy (n = 8) and those missing self-reported mobility data (n = 18) or TT (n = 154) at examination 7 were excluded, resulting in a sample size of 1445 for cross-sectional analyses (Fig. 1). To determine whether sex hormones are associated with incident and/or progression of mobility limitations, men who attended examination 8 (2005–2008) were examined on average 6.6 yr later. In the analysis of progression of mobility limitations, we excluded men who did not attend or had missing mobility data (n = 280) at examination 8 or missing covariate data (n = 1). We further excluded men reporting a mobility limitation at examination 7 (n = 53) when we examined incident mobility limitations. Hence, 1111 men were available to prospectively examine the association between circulating sex hormone levels and incident mobility limitations and 1164 men to determine the relationship between sex hormone levels and mobility limitation progression (Fig. 1).

We also determined the cross-sectional relationships between sex hormone levels and Short physical performance battery (SPPB), timed usual walk, and grip strength using a subset of men from an ancillary study to examination 7 that included the physical performance battery (n = 832, 1998–2002). Furthermore, we examined the association between sex hormone levels and the change in timed usual walk and hand grip strength among men who attended examination 8 (n = 693).

Measurement of circulating sex hormones

Serum TT, FT, and SHBG levels were measured at examination 7. TT level was measured by LC-MS/MS, as described (26–28). The sensitivity of the assay was 2 ng/dl, and interassay coefficients of variation (CV) were 7.8, 5.9, and 3.5% in samples with testosterone concentrations of 250, 500, and 1000 ng/dl, respectively. SHBG levels were measured using an immunofluorometric assay (DELFIA-Wallac, Inc., Turku, Finland). The interassay CV were 8.3, 7.9, and 10.9%, and intraassay CV were 7.3, 7.1, and 8.7%, respectively, in the low, medium, and high pools (29, 30). FT was calculated by using the law of mass action equation (31, 32). Calculated FT concentrations differ systematically from those measured by equilibrium dialysis and vary with the algorithm used for calculating FT (33).

Healthy men aged 19–40 yr enrolled in the FHS Generation 3 (children of the Offspring participants) cohort free of cardiovascular disease, cancer, diabetes, hypertension, smoking, hypercholesterolemia, and obesity (n = 456) served as the referent population to determine normative sex hormone values. For the purpose of this study, low TT and FT levels were defined as less than the 2.5th percentile for TT and FT of the referent population

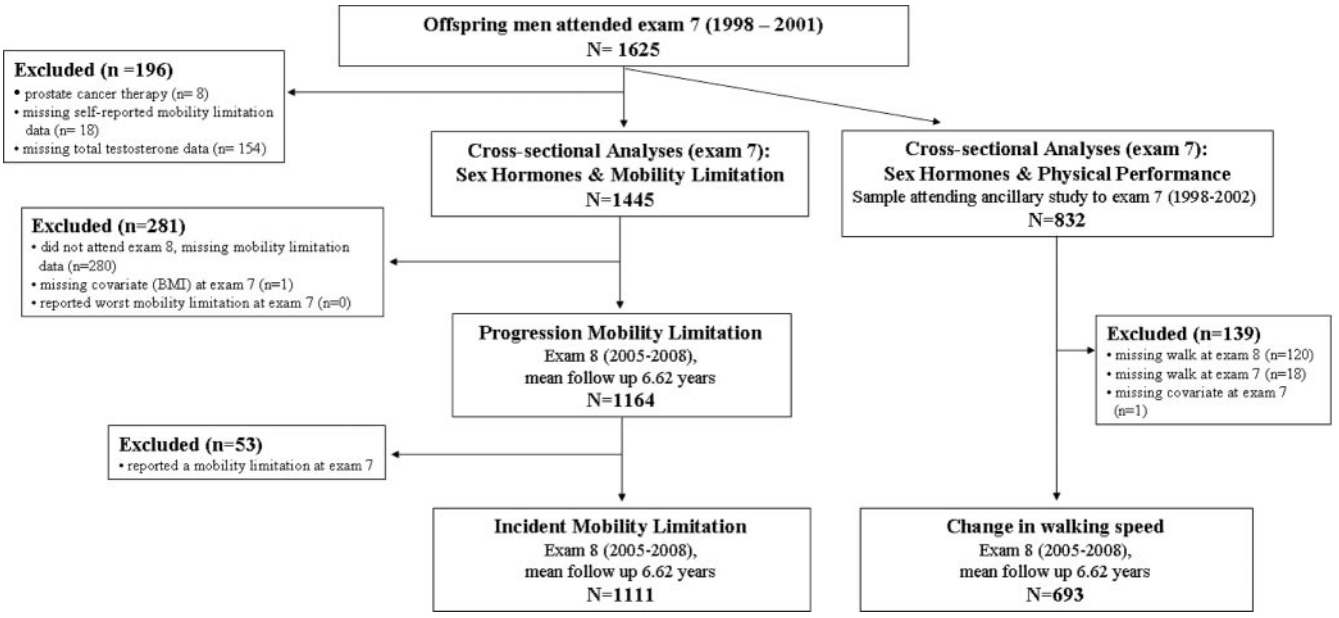


FIG. 1. Study design.

(TT <348.3 ng/dl; FT <70.0 pg/ml) and high SHBG levels as more than the 97.5th percentile of the referent sample (SHBG >81.6 nmol/liter).

Self-reported measurement of mobility limitation

At examinations 7 and 8, trained technicians queried participants about mobility limitations using a modified Rosow-Breslau questionnaire (34), which has been shown to have high test-retest reliability in other large population-based studies (35, 36). Participants were asked if they were able to 1) do heavy work around the house, like shovel snow or wash windows, walls, or floors without help; 2) walk half a mile without help (about four to six blocks); and 3) walk up and down one flight of stairs (37). At examination 7, the last item was asked as part of the Katz Activities of Daily Living scale with the following directive: during the course of a normal day, can you walk up and down one flight of stairs independently or do you need human assistance or the use of a device? Response choices included 1) no help needed, independent; 2) uses device, independent; 3) human assistance needed, minimally dependent; 4) dependent; and 5) do not do during a normal day. If the participant reported independence, he was considered able to perform the mobility task. A participant was considered to have a mobility limitation if he reported an inability to do one or more of the three items on the scale.

Subjective health

A standard single-item subjective health measure was used, "In general, how is your health now?" (examination 7) or "In general, how would you say your health is?" (examination 8). Response options at examination 7 included poor, fair, good, or excellent, and at examination 8, there was an additional response option of very good. The responses were reduced to a binary variable for analyses; 0 for good health (responses of good, very good, and excellent) and 1 for poor health (responses of poor or fair health).

Observed physical performance measures

Hand grip strength and performance-based measures of physical function were measured by trained technicians at an ancillary study to Offspring examination 7 (1998–2002). Measurements of hand grip strength and walking speed were repeated at Offspring examination 8.

SPPB

The SPPB is a validated battery that evaluates lower extremity function by measuring standing balance, gait speed, and time to rise from a chair five times (38). The standing balance measure was assigned a score ranging from 0–4, and gait speed and chair stands were assigned a score ranging from 1–4, with 4 indicating the highest level of performance. A summary performance score from 2 (worst) to 12 (best) was calculated by summing the individual scores.

Standing balance test. Participants were asked to maintain balance in three positions: feet in side by side position, feet in semi-tandem position, and feet in tandem position. For each of the three positions, participants were timed to a maximum of 10 sec. Participants were assigned a score of 0 if they were unable to hold the side-by-side standing position for 10 sec, a score of 1 if they could hold the side-by-side standing position for 10 sec but were unable to hold a semi-tandem position for 10 sec, a score of 2 if they could hold a semi-tandem position for 10 sec but were unable to hold a full-tandem position for 3 sec, a score of 3 if they could stand in a full-tandem position for 3–9 sec, or a score of 4 if they could stand in a full-tandem position for 10 sec.

Measured walk. Usual walking speed was assessed by asking the participants to walk at their usual pace over a 4-m course. Participants were allowed to use walking aids if necessary but not the assistance of another person. The test was repeated twice, and the faster of the two trials was used. Walking speed was scored as follows: less than 0.47 m/sec = 1; 0.47–0.64 m/sec = 2; 0.65–0.82 m/sec = 3; and 0.83 m/sec or faster = 4. For individuals who did not attempt or complete the walk, the value was set to the maximum value obtained by any individual.

Chair stand test. Participants were asked to stand from a sitting position in a straight-backed chair without using their arms. If

they were able to perform the task, they were asked to stand up and sit down five times, as quickly as possible. The time required to perform five chair stands was scored as follows: more than 16.6 sec = 1; 13.7–16.6 sec = 2; 11.2–13.6 sec = 3; and 11.1 sec or less = 4. If participants were unable to perform this task, then a score of 60 sec was assigned.

Hand grip strength

Grip strength was measured in both hands using an adjustable Jamar hydraulic dynamometer (Sammons Preston, Inc., Bolingbrook, IL). Participants were seated in a chair with elbow flexed at a 90° angle. Each trial consisted of a maximum squeeze for 3 sec. Three trials were performed with each hand, and the best performance in the six trials was used as the hand grip strength value.

Statistical analyses

Cross-sectional analyses

Baseline descriptive statistics (means ± SD) for continuous variables and percent for dichotomous variables were generated. Cross-sectional associations among sex hormones and binary self-reported mobility limitation and subjective health were assessed using multiple logistic regression, and multiple linear regression was used for continuous outcomes (usual walking speed, handgrip strength, and SPPB score).

Longitudinal analyses

The primary analyses employed multiple logistic regression to examine the relation between circulating sex hormone levels and 1) incident mobility limitation in men free of limitations at examination 7 and 2) decrease in subjective health between examination 7 and 8 from good or excellent to poor or fair. In secondary analyses, we examined progression of mobility limitations and decline in subjective health from examination 7 to 8, defined as a change of one or more response levels on the Rosow-Breslau scale or subjective health question (moving on Rosow-Breslau scale from 0 to 1, from 1 to 2, *etc.*). In additional analyses, we used multiple linear regression models to examine whether sex hormones measured at examination 7 were associated with change in gait speed and grip strength at examination 8 while adjusting for gait speed and grip strength at baseline (examination 7).

To account for potential confounders (variables related to outcomes that might affect the strength of association), all models were adjusted for age, body mass index (BMI), smoking, and comorbidities (cardiovascular disease and cancer) at examination 7. However, the univariate association of TT and FT levels with the Framingham physical activity index, a measure of physical activity, was either very weak (Pearson's correlation coefficient for TT = 0.07; $P = 0.008$) or not significant (Pearson correlation coefficient for FT = 0.01; $P = 0.6867$). Therefore, the analyses were not adjusted for physical activity index.

Furthermore, to examine the potential threshold effect, where hormone concentrations below a certain level relate to risk of poor outcomes, both cross-sectional and longitudinal analyses were repeated, defining low levels of TT and FT and high levels of SHBG based on the 2.5th percentile cutoff obtained from the Generation 3 healthy reference sample. Statistical significance level was set at two-sided $P < 0.05$.

Results

Demographic data

The baseline characteristics of men in our study sample with sex hormone and mobility data are shown in Table 1. The men in our sample were on average 61.0 yr at baseline with mean TT, FT, and SHBG levels of 583 ± 227 ng/dl, 86 ± 32 pg/ml, and 58 ± 27 nmol/liter, respectively, and 15.4% had low TT, 31.6% had low FT, and 15.5% had high SHBG levels. The proportion of men with self-reported mobility limitation at baseline was 6.4 and 7.1% of men reported poor subjective health. In the sample of men with physical performance data, the mean SPPB score was 10.9, usual walking speed 1.25 m/sec, and hand grip strength 42.4 kg. Thirteen percent (n = 144) of the sample reported an occurrence of mobility limitation at examination 8, and 14% (n = 163) reported a progression of their mobility limitation from examination 7 to 8.

Cross-sectional relation between sex hormones and mobility and physical performance

The cross-sectional associations between circulating levels of sex hormones and self-reported mobility limitations and subjective health at examination 7 are presented in Table 2. TT and SHBG were not significantly associated with mobility limitation or subjective health. FT levels were not significantly associated with mobility but were associated with subjective health. As FT increased, the chances of reporting poor subjective health decreased; 1 SD increase in FT was associated with a 28% decrease in the odds of reporting poor subjective health [multivariable-adjusted odds ratio (OR) = 0.72; 95% confidence interval (CI) = 0.56–0.94]. Compared with men with normal FT levels, men with low FT levels had an increased risk of reporting poor subjective health (OR = 1.61; 95% CI = 1.02–2.55). The cross-sectional associations between circulating levels of sex hormones and physical performance measures at baseline examination 7 are presented in Table 3. TT and SHBG were not significantly associated with any of the physical performance measures. FT levels were significantly associated with SPPB score and usual walking speed. As FT increased, SPPB score and usual walking speed increased as well; each SD increase in FT was associated with a 0.13-U increase in SPPB score ($P = 0.008$) and 0.02 m/sec increase in usual walking speed ($P = 0.048$). Men with low FT were also more likely to have lower grip strength (adjusted mean difference between men with low and high FT = −2.01; 95% CI = −3.95 to −0.07) than those with normal FT. Low TT and high SHBG were not significantly associated with any mobility or physical performance measure.

TABLE 1. Baseline characteristics at examination 7

Characteristic	Cross-sectional analyses: Men with mobility limitation data (n = 1445)	Cross-sectional analyses: Men with physical performance data (n = 832)	Longitudinal analyses: men with incident mobility limitation data (n = 1111)
Age (yr)	61.0 (9.5)	61.6 (9.3)	59.6 (9.0)
Smoking (%)	12.7	11.5	11.8
Alcohol consumption, drinks/wk (%)			
None	27.3	28.9	25.8
1–14	69.1	67.4	71.1
>14	3.6	3.7	3.1
BMI (kg/m^2)	28.8 (4.5)	28.8 (4.5)	28.7 (4.5)
Prevalent cardiovascular disease (%)	17.6	17.3	13.1
Cancer (%)	9.7	10.8	7.7
TT (ng/dl)	583.5 (226.5)	584.4 (229.4)	589.9 (228.7)
FT (pg/ml)	86.1 (31.8)	86.2 (32.0)	88.6 (31.4)
SHBG (nmol/liter)	58.2 (26.7)	58.2 (26.3)	56.4 (25.5)
Men with low TT (%)[a]	15.4	15.1	15.0
Men with low FT (%)	31.6	31.3	27.7
Men with high SHBG (%)	15.5	15.5	14.0
Self-reported mobility limitation (%)	6.4	5.8	0.0
Poor subjective health (%)	7.1	6.4	4.9
SPPB summary score		10.9 (1.30)	11.0 (1.10)
Usual walking speed (m/sec)		1.25 (0.30)	1.28 (0.29)
Grip strength (kg)		42.4 (12.5)	43.5 (12.0)

Values are mean (SD) for continuous variables and percentages for dichotomous characteristics. To convert TT to SI units (nanomoles per liter), multiply TT concentrations in nanograms per deciliter by 0.0347. To convert FT to SI units (picomoles per liter), multiply FT concentration in picograms per milliliter by 3.47.

[a] Sex hormones were defined as low or high *vs.* normal using healthy reference sample of FHS Generation 3 men. TT and FT levels below the 2.5th percentile of the referent sample (TT <348.3 ng/dl; FT <70.0 pg/ml) were deemed low, and SHBG levels above the 97.5th percentile of the referent sample (SHBG, 81.6 nmol/liter) were deemed high levels.

Longitudinal relation between sex hormones and mobility and physical performance

The results of our primary analyses of the impact of hormones on development of mobility limitations and decline in subjective health are presented in Table 4. Baseline low FT was a significant predictor of incident mobility limitation (Fig. 2). As FT increased by 1 SD, the risk of developing mobility limitation decreased by 22% (OR =

TABLE 2. Cross-sectional associations between baseline circulating sex hormone levels and self-reported mobility limitation at examination 7 (n = 1445)

	Multivariable logistic regression: Mobility limitation OR (95% CI)	*P*	Subjective health OR (95% CI)	*P*
Continuous hormone levels				
TT	0.89 (0.70–1.13)	0.34	0.80 (0.63–1.01)	0.06
FT	0.80 (0.61–1.03)	0.09	0.72 (0.56–0.94)	0.01
SHBG	1.13 (0.91–1.40)	0.28	1.05 (0.85–1.31)	0.64
Dichotomized sex hormone levels				
Low TT	1.12 (0.62–2.02)	0.70	1.09 (0.62–1.89)	0.77
Low FT	1.29 (0.80–2.08)	0.30	1.61 (1.02–2.55)	0.04
High SHBG	1.31 (0.75–2.28)	0.34	1.20 (0.69–2.09)	0.51

Continuous hormone levels and OR values are for 1 SD change in sex hormone levels. All models were adjusted for age, BMI, smoking, and comorbidities (cancer and cardiovascular disease) at examination 7. Sex hormones were defined as low or high *vs.* normal using healthy reference sample of FHS Generation 3 men. Low TT and FT levels were those below the 2.5th percentile of the referent sample (TT <348.3 ng/dl; FT <70.0 pg/ml), and SHBG levels above the 97.5th percentile of the referent sample (SHBG, 81.6 nmol/liter) represented high SHBG levels. Low FT levels (<70 pg/ml) were associated with increased risk (OR = 1.61) of poor subjective health; each SD increase in FT level was associated with a 28% decease (OR = 0.72) in risk of reporting poor subjective health.

TABLE 3. Cross-sectional associations between baseline circulating sex hormone levels and physical performance at examination 7 (n = 832)

	Multivariable linear regression					
	SPPB score		Usual walking speed (m/sec)		Grip strength (kg)	
	β (95% CI)	P	β (95% CI)	P	β (95% CI)	P
Continuous hormone levels						
TT	0.06 (−0.03–0.15)	0.18	0.02 (−0.006–0.04)	0.15	0.35 (−0.55–1.24)	0.45
FT	0.13 (0.03–0.22)	0.008	0.02 (0.0001–0.04)	0.048	0.57 (−0.33–1.48)	0.21
SHBG	−0.08 (−0.17–0.02)	0.13	−0.01 (−0.04–0.009)	0.23	−0.41 (−1.38–0.56)	0.41
Dichotomized sex hormone levels						
Low TT	−0.14 (−0.40–0.11)	0.27	−0.01 (−0.07–0.05)	0.65	−1.73 (−4.16–0.70)	0.16
Low FT	−0.13 (−0.34–0.07)	0.20	−0.03 (−0.08–0.01)	0.16	−2.01 (−3.95–0.07)	0.04
High SHBG	−0.08 (−0.33–0.18)	0.55	−0.05 (−0.11–0.007)	0.09	0.38 (−2.13–2.88)	0.77

Continuous hormone levels and β−values are for 1 SD change in sex hormone levels. All models were adjusted for age, BMI, smoking, and comorbidities (cancer and cardiovascular disease) at examination 7. Sex hormones were defined as low or high *vs.* normal using healthy reference sample of FHS Generation 3 men. Low TT and FT levels were those below the 2.5th percentile of the referent sample (TT <348.3 ng/dl; FT <70.0 pg/ml), and SHBG levels above the 97.5th percentile of the referent sample (SHBG, 81.6 nmol/liter) represented high SHBG levels. Low FT levels (<70 pg/ml) were associated with decreased grip strength. Each SD increase in FT level was associated with 0.13 U increase in SPPB score and 0.02 m/sec increase in usual walking speed.

0.78; 95% CI = 0.62–0.97). FT levels were also significantly associated with progression of mobility limitation (OR = 0.75; 95% CI = 0.60–0.93). Thus, men were 25% less likely to report worsening mobility limitation for each SD increase in circulating FT. No significant relationships were observed between FT and subjective health in the longitudinal analysis. TT, FT, and SHBG levels were not significantly associated with change in usual walking speed or handgrip strength from examination 7 to 8 (data not shown).

Men with low FT were 57% more likely to develop mobility limitation on follow-up (OR = 1.57; 95% CI = 1.06–2.32) and 68% more likely to experience worsening of mobility limitation (OR = 1.68; 95% CI = 1.16–2.45) compared with men with normal FT (Table 4 and Figure 2). Low TT and high SHBG were not associated with either incident mobility limitation or progression of mobility limitation.

Discussion

In our community-based sample of men, higher baseline FT levels (both continuous and threshold values) were significantly associated with lower odds of an incident mobility limitation. Baseline FT was also a significant cor-

TABLE 4. Longitudinal relations between baseline circulating sex hormone levels and incident mobility limitation: multivariable logistic regression (n = 1111)

	Mobility limitation		Subjective health	
	OR (95% CI)	P	OR (95% CI)	P
Continuous hormone levels				
TT	0.90 (0.74–1.09)	0.28	1.09 (0.80–1.47)	0.59
FT	0.78 (0.62−0.97)	0.03	0.87 (0.62–1.22)	0.42
SHBG	1.11 (0.91–1.34)	0.31	1.29 (0.97–1.72)	0.08
Dichotomized sex hormone levels				
Low TT	1.46 (0.91–2.34)	0.12	0.48 (0.19–1.24)	0.13
Low FT	1.57 (1.06–2.32)	0.03	1.23 (0.64–2.38)	0.53
High SHBG	1.29 (0.80–2.10)	0.30	1.59 (0.74–3.43)	0.24

OR values are for 1 SD change in hormone levels. For incidence, the sample excludes subjects at exam 7 with mobility limitation and poor subjective health. For progression, the sample excludes subjects at exam 7 with worst response choice for mobility limitation and subjective health. All models were adjusted for age, BMI, smoking, and comorbidities (cancer and cardiovascular disease). Sex hormones were defined as low or high *vs.* normal using healthy reference sample of FHS Generation 3 men. Low TT and FT levels were those below the 2.5th percentile of the referent sample (TT <348.3 ng/dl; FT <70.0 pg/ml), and SHBG levels above the 97.5th percentile of the referent sample (SHBG, 81.6 nmol/liter) represented high SHBG levels. Low FT levels (<70 pg/ml) at examination 7 were associated with increased risk (OR = 1.57) of developing mobility limitation at examination 8. Each SD increase in FT level at examination 7 was associated with a 22% decrease (OR = 0.78) in risk of reporting mobility limitation at examination 8.

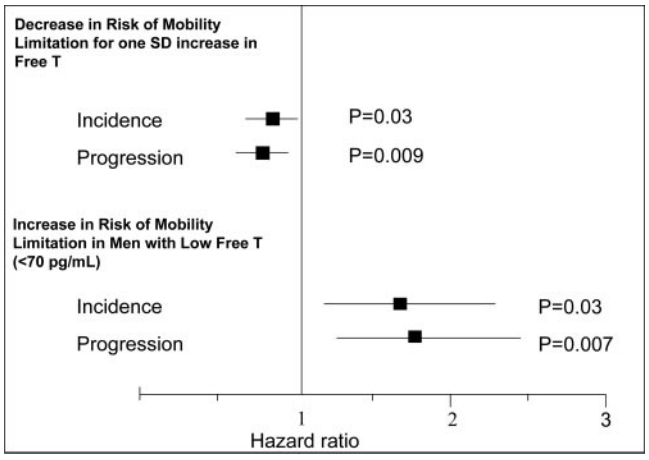


FIG. 2. Longitudinal analyses of incident mobility limitation. Continuous FT level hazard ratios are for 1 SD increase in hormone levels, adjusting for age, BMI, smoking, and comorbidities (cardiovascular disease and cancer). As shown in the *upper panel*, each SD increase in FT level was associated with 22% (OR = 0.78; 95% CI = 0.62–0.97) decrease in the risk of developing mobility limitation and 25% decrease in the risk of worsening mobility limitation (progression). The *lower panel* shows the association of low FT (<2.5th percentile (<70.0 pg/ml)) at baseline examination 7 with the risk of developing (incident) mobility limitation at examination 8 or of reporting worsening mobility limitation (progression) at examination 8. The FT hazard ratios were adjusted for age, BMI, smoking, and comorbidities. The *squares* indicate point estimates for hormones, and the *lines* indicate 95% CI.

relate of progression of mobility limitation, consistent with the incidence findings. Furthermore, FT was positively associated with faster baseline usual walking speed and SPPB score, a valid measure of lower extremity function and an important determinant of mobility. Thus, baseline FT is a significant correlate of both self-reported and performance-based measures of mobility. TT and SHBG were not associated with any of the mobility or physical performance measures.

According to the free hormone hypothesis, FT, representing the unbound hormone, is considered the biologically active fraction of testosterone. Although bioavailable testosterone has been reported to be associated with self-reported mobility limitation, muscle strength, and physical performance measures (39), we did not analyze bioavailable testosterone because it is a calculated multiple of FT. However, recent data suggest that SHBG-bound testosterone may be internalized through endocytic pits after binding to the megalin receptor and may also be biologically relevant (40). Indeed, in the Massachusetts Male Ageing Study, SHBG, rather than TT or FT, was associated with frailty (41). Our data support the free hormone hypothesis and suggest that FT may mediate most of the effects of testosterone on physical function measures because we did not find any relationship between SHBG and TT with mobility or physical performance measures.

FT levels were associated with subjective health in cross-sectional but not longitudinal analyses. Although testosterone may not be causally related to subjective health, it is possible that factors that contribute to poor subjective health such as comorbid conditions may also lower testosterone levels. It is possible that individuals whose health deteriorated between examinations 7 and 8 did not return for follow-up. Of the men that did not return for examination 8, or had missing incident mobility limitation at examination 8, 11.8% had reported poor subjective health at examination 7 compared with 7.1% of men in the total sample. This may also explain why fewer men reported poor subjective health at examination 8. Therefore, the positive longitudinal associations between FT and mobility limitations and their progression are all the more remarkable.

The observed association between FT and mobility measures has biological plausibility. Testosterone is an important determinant of skeletal muscle mass (17) and increases muscle mass by promoting myogenic differentiation of multipotent mesenchymal stem cells (42, 43) and by stimulating muscle protein synthesis (44, 45). Testosterone administration also increases maximal voluntary strength and power in men (29, 30, 46). However, the association of testosterone with physical function measures in epidemiological studies has been inconsistent. Although some studies have found testosterone levels to be related to self-reported (19) as well as performance-based measures of physical function (18, 21, 47), frailty (48), and falls (39), a recent prospective analysis of two cohorts did not find any significant association of either TT or FT with decline in physical function or muscle strength (20). The effects of testosterone therapy on physical function measures in randomized testosterone trials have been heterogeneous. Some trials have reported improvements in gait speed, stair climbing power, and composite measures of physical function (49, 50), whereas others failed to find significant effects (51–53). However, older men included in the first-generation testosterone trials were not uniformly hypogonadal (51–53). Also, most of the studies included healthy older men without functional limitations and used tests of physical function that had low ceiling (54). Finally, testosterone doses in some trials were small and did not significantly raise serum testosterone (51, 52). We have shown that testosterone administration in young and older men is associated with dose-dependent increments in skeletal muscle mass and maximal voluntary strength (30, 55, 56).

We observed that circulating FT levels were significantly associated with both SPPB scores and walking speed. Each SD increase in FT was associated with a 0.13-U increase in SPPB score and 0.02-m/sec in usual walking speed. Perera *et al.* (57) have deemed a 0.5-U change in SPPB score and a 0.05-m/sec change in gait speed to be

clinically important changes. Thus, testosterone levels have a small but significant effect on these measures of physical performance. Indeed, testosterone is only one of many physiological processes that regulate complex functions such as walking, although it is an important remediable factor and, therefore, the subject of current investigation.

Our study has significant strengths. First, the FHS cohort included community-dwelling men over a wider age range than has been included in some other studies that were focused mostly on older men. The longitudinal design of our analyses lends strength to our inferences. We adjusted our analyses for potential confounders, including age, BMI, smoking, and comorbidities. This is the first population study to evaluate the relationship between mobility and physical function with TT levels measured by LC-MS/MS, widely considered the gold standard for the measurement of testosterone levels (23). We defined reference ranges of testosterone levels by using healthy young men, age 19–40 yr, and evaluated the relationship between androgens and mobility based on these population-based thresholds.

Our study also has some limitations. First, epidemiological studies including longitudinal studies can define associations but not causality. Second, our study population was white, and therefore, our findings may not be generalizable to other race/ethnicities. Also, of the 1445 men who were evaluated for mobility limitation at examination 7, 280 did not return for examination 8 or had missing mobility limitation data at examination 8. Importantly, men who did not return for examination 8 had a higher frequency of mobility limitation, lower SPPB score, and slower walking speed at baseline than those who did return for examination 8. Thus, it is possible that some of the subjects with poor health whose health deteriorated did not return for follow-up, thus diluting the observed effects. Therefore, our longitudinal analyses likely represent a conservative estimate of the association between FT levels and mobility limitation and walking speed. We did not have sex hormones measured at examination 8 to evaluate the correlation between the change in testosterone levels and incident mobility. We did not measure estradiol levels and were unable to dissect out the possible role of aromatization on these outcomes. Serum testosterone levels are affected by pulsatile, diurnal, and circannual rhythms, and single samples ignore rhythmic hormone secretion. Our analyses show that single early morning testosterone levels, obtained in a manner similar to that by physicians in real practice, were associated with mobility limitation and some other measures of physical function. Therefore, even though our models did not factor in the complexities of biological rhythms, they are in concordance with the need of practitioners to depend on conveniently obtained single samples. Finally, the Framingham cohort was younger and healthier than some other epidemiological studies, resulting in fewer events and lower rates of worsening of physical function; this may have reduced the statistical power. We had 0.685 power to detect an association of similar magnitude for TT to that obtained from FT at an α of 0.05. The follow-up of these men over a still longer period of time resulting in potentially more events could increase the power.

These data have clinical implications. Mobility is one of the most important physical functions, essential for independent living. Our data show that men with low FT had a 57% greater risk of developing a mobility limitation and a 68% higher risk of deterioration in their mobility. Whether this risk can be reduced with exogenous testosterone therapy in older men with mobility limitation and low free testosterone levels would need to be determined by a randomized clinical trial.

Acknowledgments

Address all correspondence and requests for reprints to: Shalender Bhasin, M.D., School of Medicine, Section of Endocrinology, Diabetes, and Nutrition, 670 Albany Street, Second Floor, Boston, Massachusetts 02118-2393. E-mail: Bhasin@bu.edu.

This work was supported by National Institute on Aging, 1RO1AG31206; Boston Claude D. Pepper Older Americans Independence Center, 5P30 AG31679; and NIH grants AG029451-01A2 and 2R01 AG16495. 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.

Disclosure Summary: S.B. received research support from the Centers for Disease Control Foundation.

References

1. **Shumway-Cook A, Ciol MA, Yorkston KM, Hoffman JM, Chan L** 2005 Mobility limitations in the Medicare population: prevalence and sociodemographic and clinical correlates. J Am Geriatr Soc 53: 1217–1221
2. **Rubenstein LZ, Powers CM, MacLean CH** 2001 Quality indicators for the management and prevention of falls and mobility problems in vulnerable elders. Ann Intern Med 135:686–693
3. **Guralnik JM, Ferrucci L, Simonsick EM, Salive ME, Wallace RB** 1995 Lower-extremity function in persons over the age of 70 years as a predictor of subsequent disability. N Engl J Med 332:556–561
4. **von Bonsdorff M, Rantanen T, Laukkanen P, Suutama T, Heikkinen E** 2006 Mobility limitations and cognitive deficits as predictors of institutionalization among community-dwelling older people. Gerontology 52:359–365
5. **Groessl EJ, Kaplan RM, Rejeski WJ, Katula JA, King AC, Frierson G, Glynn NW, Hsu FC, Walkup M, Pahor M** 2007 Health-related

quality of life in older adults at risk for disability. Am J Prev Med 33:214–218
6. **Metter EJ, Schrager M, Ferrucci L, Talbot LA** 2005 Evaluation of movement speed and reaction time as predictors of all-cause mortality in men. J Gerontol A Biol Sci Med Sci 60:840–846
7. **Hurley BF** 1995 Age, gender, and muscular strength. J Gerontol A Biol Sci Med Sci 50(Spec No):41–44
8. **Lauretani F, Russo CR, Bandinelli S, Bartali B, Cavazzini C, Di Iorio A, Corsi AM, Rantanen T, Guralnik JM, Ferrucci L** 2003 Age-associated changes in skeletal muscles and their effect on mobility: an operational diagnosis of sarcopenia. J Appl Physiol 95:1851–1860
9. **Feldman HA, Longcope C, Derby CA, Johannes CB, Araujo AB, Coviello AD, Bremner WJ, McKinlay JB** 2002 Age trends in the level of serum testosterone and other hormones in middle-aged men: longitudinal results from the Massachusetts male aging study. J Clin Endocrinol Metab 87:589–598
10. **Harman SM, Metter EJ, Tobin JD, Pearson J, Blackman MR** 2001 Longitudinal effects of aging on serum total and free testosterone levels in healthy men. Baltimore Longitudinal Study of Aging. J Clin Endocrinol Metab 86:724–731
11. **Morley JE, Kaiser FE, Perry 3rd HM, Patrick P, Morley PM, Stauber PM, Vellas B, Baumgartner RN, Garry PJ** 1997 Longitudinal changes in testosterone, luteinizing hormone, and follicle-stimulating hormone in healthy older men. Metabolism 46:410–413
12. **Simon D, Preziosi P, Barrett-Connor E, Roger M, Saint-Paul M, Nahoul K, Papoz L** 1992 The influence of aging on plasma sex hormones in men: the Telecom Study. Am J Epidemiol 135:783–791
13. **Wu FC, Tajar A, Pye SR, Silman AJ, Finn JD, O'Neill TW, Bartfai G, Casanueva F, Forti G, Giwercman A, Huhtaniemi IT, Kula K, Punab M, Boonen S, Vanderschueren D** 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
14. **Zmuda JM, Cauley JA, Kriska A, Glynn NW, Gutai JP, Kuller LH** 1997 Longitudinal relation between endogenous testosterone and cardiovascular disease risk factors in middle-aged men. A 13-year follow-up of former Multiple Risk Factor Intervention Trial participants. Am J Epidemiol 146:609–617
15. **Muller M, den Tonkelaar I, Thijssen JH, Grobbee DE, van der Schouw YT** 2003 Endogenous sex hormones in men aged 40–80 years. Eur J Endocrinol 149:583–589
16. **Roy TA, Blackman MR, Harman SM, Tobin JD, Schrager M, Metter EJ** 2002 Interrelationships of serum testosterone and free testosterone index with FFM and strength in aging men. Am J Physiol Endocrinol Metab 283:E284–E294
17. **Baumgartner RN, Waters DL, Gallagher D, Morley JE, Garry PJ** 1999 Predictors of skeletal muscle mass in elderly men and women. Mech Ageing Dev 107:123–136
18. **Schaap LA, Pluijm SM, Smit JH, van Schoor NM, Visser M, Gooren LJ, Lips P** 2005 The association of sex hormone levels with poor mobility, low muscle strength and incidence of falls among older men and women. Clin Endocrinol (Oxf) 63:152–160
19. **O'Donnell AB, Travison TG, Harris SS, Tenover JL, McKinlay JB** 2006 Testosterone, dehydroepiandrosterone, and physical performance in older men: results from the Massachusetts Male Aging Study. J Clin Endocrinol Metab 91:425–431
20. **Schaap LA, Pluijm SM, Deeg DJ, Penninx BW, Nicklas BJ, Lips P, Harris TB, Newman AB, Kritchevsky SB, Cauley JA, Goodpaster BH, Tylavsky FA, Yaffe K, Visser M** 2008 Low testosterone levels and decline in physical performance and muscle strength in older men: findings from two prospective cohort studies. Clin Endocrinol (Oxf) 68:42–50
21. **Araujo AB, Travison TG, Bhasin S, Esche GR, Williams RE, Clark RV, McKinlay JB** 2008 Association between testosterone and estradiol and age-related decline in physical function in a diverse sample of men. J Am Geriatr Soc 56:2000–2008
22. **Bhasin S, Zhang A, Coviello A, Jasuja R, Ulloor J, Singh R, Vesper H, Vasan RS** 2008 The impact of assay quality and reference ranges on clinical decision making in the diagnosis of androgen disorders. Steroids 73:1311–1317
23. **Rosner W, Auchus RJ, Azziz R, Sluss PM, Raff H** 2007 Position statement: Utility, limitations, and pitfalls in measuring testosterone: an Endocrine Society position statement. J Clin Endocrinol Metab 92:405–413
24. **Dawber TR, Meadors GF, Moore Jr FE** 1951 Epidemiological approaches to heart disease: the Framingham Study. Am J Public Health Nations Health 41:279–281
25. **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
26. **Sattler FR, Castaneda-Sceppa C, Binder EF, Schroeder ET, Wang Y, Bhasin S, Kawakubo M, Stewart Y, Yarasheski KE, Ulloor J, Colletti P, Roubenoff R, Azen SP** 2009 Testosterone and growth hormone improve body composition and muscle performance in older men. J Clin Endocrinol Metab 94:1991–2001
27. **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
28. **Vesper HW, Botelho JC, Shacklady C, Smith A, Myers GL** 2008 CDC project on standardizing steroid hormone measurements. Steroids 73:1286–1292
29. **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
30. **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
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. **Vermeulen A, Verdonck L, Kaufman JM** 1999 A critical evaluation of simple methods for the estimation of free testosterone in serum. J Clin Endocrinol Metab 84:3666–3672
33. **Sartorius G, Ly LP, Sikaris K, McLachlan R, Handelsman DJ** 2009 Predictive accuracy and sources of variability in calculated free testosterone estimates. Ann Clin Biochem 46:137–143
34. **Rosow I, Breslau N** 1966 A Guttman health scale for the aged. J Gerontol 21:556–559
35. **Beckett LA, Brock DB, Lemke JH, Mendes de Leon CF, Guralnik JM, Fillenbaum GG, Branch LG, Wetle TT, Evans DA** 1996 Analysis of change in self-reported physical function among older persons in four population studies. Am J Epidemiol 143:766–778
36. **Crawford SL, Jette AM, Tennstedt SL** 1997 Test-retest reliability of self-reported disability measures in older adults. J Am Geriatr Soc 45:338–341
37. **Brorsson B, Asberg KH** 1984 Katz index of independence in ADL. Reliability and validity in short-term care. Scand J Rehabil Med 16:125–132
38. **Guralnik JM, Simonsick EM, Ferrucci L, Glynn RJ, Berkman LF, Blazer DG, Scherr PA, Wallace RB** 1994 A short physical performance battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission. J Gerontol 49:M85–M94
39. **Orwoll E, Lambert LC, Marshall LM, Blank J, Barrett-Connor E, Cauley J, Ensrud K, Cummings SR** 2006 Endogenous testosterone levels, physical performance, and fall risk in older men. Arch Intern Med 166:2124–2131
40. **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
41. **Mohr BA, Bhasin S, Kupelian V, Araujo AB, O'Donnell AB, McKinlay JB** 2007 Testosterone, sex hormone-binding globulin, and frailty in older men. J Am Geriatr Soc 55:548–555
42. **Singh R, Artaza JN, Taylor WE, Gonzalez-Cadavid NF, Bhasin S** 2003 Androgens stimulate myogenic differentiation and inhibit adipogenesis in C3H 10T1/2 pluripotent cells through an androgen receptor-mediated pathway. Endocrinology 144:5081–5088
43. **Singh R, Bhasin S, Braga M, Artaza JN, Pervin S, Taylor WE, Krishnan V, Sinha SK, Rajavashisth TB, Jasuja R** 2009 Regulation of myogenic differentiation by androgens: cross talk between androgen receptor/β-catenin and follistatin/transforming growth factor-β signaling pathways. Endocrinology 150:1259–1268
44. **Brodsky IG, Balagopal P, Nair KS** 1996 Effects of testosterone replacement on muscle mass and muscle protein synthesis in hypogonadal men: a clinical research center study. J Clin Endocrinol Metab 81:3469–3475
45. **Ferrando AA, Sheffield-Moore M, Paddon-Jones D, Wolfe RR, Urban RJ** 2003 Differential anabolic effects of testosterone and amino acid feeding in older men. J Clin Endocrinol Metab 88:358–362
46. **Bhasin S, Storer TW, Berman N, Callegari C, Clevenger B, Phillips J, Bunnell TJ, Tricker R, Shirazi A, Casaburi R** 1996 The effects of supraphysiologic doses of testosterone on muscle size and strength in normal men. N Engl J Med 335:1–7
47. **Szulc P, Claustrat B, Marchand F, Delmas PD** 2003 Increased risk of falls and increased bone resorption in elderly men with partial androgen deficiency: the MINOS study. J Clin Endocrinol Metab 88:5240–5247
48. **Cawthon PM, Ensrud KE, Laughlin GA, Cauley JA, Dam TT, Barrett-Connor E, Fink HA, Hoffman AR, Lau E, Lane NE, Stefanick ML, Cummings SR, Orwoll ES** 2009 Sex hormones and frailty in older men: the osteoporotic fractures in men (MrOS) study. J Clin Endocrinol Metab 94:3806–3815
49. **Brill KT, Weltman AL, Gentili A, Patrie JT, Fryburg DA, Hanks JB, Urban RJ, Veldhuis JD** 2002 Single and combined effects of growth hormone and testosterone administration on measures of body composition, physical performance, mood, sexual function, bone turnover, and muscle gene expression in healthy older men. J Clin Endocrinol Metab 87:5649–5657
50. **Page ST, Amory JK, Bowman FD, Anawalt BD, Matsumoto AM, Bremner WJ, Tenover JL** 2005 Exogenous testosterone (T) alone or with finasteride increases physical performance, grip strength, and lean body mass in older men with low serum T. J Clin Endocrinol Metab 90:1502–1510
51. **Emmelot-Vonk MH, Verhaar HJ, Nakhai Pour HR, Aleman A, Lock TM, Bosch JL, Grobbee DE, van der Schouw YT** 2008 Effect of testosterone supplementation on functional mobility, cognition, and other parameters in older men: a randomized controlled trial. JAMA 299:39–52
52. **Nair KS, Rizza RA, O'Brien P, Dhatariya K, Short KR, Nehra A, Vittone JL, Klee GG, Basu A, Basu R, Cobelli C, Toffolo G, Dalla Man C, Tindall DJ, Melton 3rd LJ, Smith GE, Khosla S, Jensen MD** 2006 DHEA in elderly women and DHEA or testosterone in elderly men. N Engl J Med 355:1647–1659
53. **Snyder PJ, Peachey H, Hannoush P, Berlin JA, Loh L, Lenrow DA, Holmes JH, Dlewati A, Santanna J, Rosen CJ, Strom BL** 1999 Effect of testosterone treatment on body composition and muscle strength in men over 65 years of age. J Clin Endocrinol Metab 84:2647–2653
54. **LeBrasseur NK, Bhasin S, Miciek R, Storer TW** 2008 Tests of muscle strength and physical function: reliability and discrimination of performance in younger and older men and older men with mobility limitations. J Am Geriatr Soc 56:2118–2123
55. **Bhasin S, Woodhouse L, Casaburi R, Singh AB, Bhasin D, Berman N, Chen X, Yarasheski KE, Magliano L, Dzekov C, Dzekov J, Bross R, Phillips J, Sinha-Hikim I, Shen R, Storer TW** 2001 Testosterone dose-response relationships in healthy young men. Am J Physiol Endocrinol Metab 281:E1172–E1181
56. **Storer TW, Magliano L, Woodhouse L, Lee ML, Dzekov C, Dzekov J, Casaburi R, Bhasin S** 2003 Testosterone dose-dependently increases maximal voluntary strength and leg power, but does not affect fatigability or specific tension. J Clin Endocrinol Metab 88:1478–1485
57. **Perera S, Mody SH, Woodman RC, Studenski SA** 2006 Meaningful change and responsiveness in common physical performance measures in older adults. J Am Geriatr Soc 54:743–749

Self-Reported Willingness to Have Cancer Screening and the Effects of Sociodemographic Factors

Nancy R. Kressin, PhD; Meredith Manze, MPH; Stefanie L. Russell, DDS, MPH, PhD; Ralph V. Katz, DMD, MPH, PhD; Cristina Claudio, PhD; B. Lee Green, PhD; Min Qi Wang, PhD

Disclaimer: The views expressed in this article are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs.

Funding/Support: This study was supported by U54 DE 14257, and a VA Research Career Scientist award from the VA Health Services Research and Development Service to Dr Kressin.

Background: The relative effects of race/ethnicity and other sociodemographic factors, compared to those of attitudes and beliefs on willingness to have cancer screening, are not well understood.

Methods: We conducted telephone interviews with 1148 adults (31% African American, 27% Puerto Rican American, 43% white) from 3 cities in mainland United States and Puerto Rico. Respondents reported their sociodemographic characteristics, attitudes about barriers and facilitators of cancer screening, and willingness to have cancer screening under 4 scenarios: when done in the community vs one's doctor's office, and whether or not one had symptoms.

Results: Racial/ethnic minority status, age, and lower income were frequently associated with increased willingness to have cancer screening, even after including attitudes and beliefs about screening. Having screening nearby was important for community screening, and anticipation of embarrassment from screening for when there were no cancer symptoms. Associations varied across 4 screening scenarios, with the fewest predictors for screening by one's doctor when there were symptoms.

Conclusions: Sociodemographic characteristics not only were related to willingness to have cancer screenings in almost all cases, but were generally much stronger factors than attitudinal barriers and facilitators. Cancer screening campaigns should affect attitudinal change where possible, but should also recognize that targeting screening to specific population groups may be necessary.

Keywords: cancer screening ■ knowledge, attitudes, and beliefs ■ minority health

J Natl Med Assoc. 2010;102:219-227

Author Affiliations: Section of General Internal Medicine, Health Care Disparities Research Program, Boston University School of Medicine, Boston, Massachusetts (Dr Kressin and Ms Manze); VA Boston Healthcare System, Jamaica Plain, Massachusetts (Dr Kressin); Center for Health Quality, Outcomes & Economic Research, VAMC, Bedford, Massachusetts (Dr Kressin); Department of Epidemiology and Health Promotion, NYU College of Dentistry, New York University, New York, New York (Drs Russell and Katz); Community Dentistry Section, School of Dentistry, University of Puerto Rico, San Juan, Puerto Rico (Dr Claudio); Office of Institutional Diversity, and Health Outcomes and Behavior, H. Lee Moffitt Cancer Center and Research Institute, Tampa, Florida (Dr Green); Department of Public and Community Health, University of Maryland, College Park, MD (Dr Wang).

Correspondence: Nancy R. Kressin, PhD, Section of General Internal Medicine, Health/Care Disparities Research Program, Boston University School of Medicine, 801 Massachusetts Ave, #2090, Boston, MA 02118 (nkressin@bu.edu).

INTRODUCTION

Cancer screening is crucial to identifying cancer in its early stages, when the disease is more amenable to treatment or cure. Screening rates among racial and ethnic minorities vary compared to whites, with minorities having lower screening rates for certain types of cancer, such as cervical and colorectal cancer.[1-3] Individuals with less education and income receive cancer screening less often than do those with higher levels of each.[4] These differential rates in screening may lead to disparities in cancer-related mortality[5-8] and highlight the need to understand the reasons for cancer screening disparities.[9-11] Several studies have examined the association of race/ethnicity with potential barriers to screening utilization, such as access to care through health insurance; relationship with health care providers; provider recommendation of screening; patient knowledge, and attitudes and beliefs about screening.[12-25] Often, these studies, many of them qualitative in nature, have focused on understanding the dynamics of these issues among a single racial/ethnic group, and they emphasized the importance of identifying the health beliefs of certain groups to help tailor interventions and understand their decisions for cancer screening.

However, the relative effects of race/ethnicity on willingness to have cancer screening, compared to other sociodemographic factors (including education, income,

and employment status), as well as to attitudes and beliefs about screening, are not well understood.[26] Although race has been associated with numerous negative predictors of cancer screening, it may be that many of the factors associated with race, rather than—or in addition to—race itself, are important driving forces behind such associations, as has been found with regard to health care utilization in general.[27,28] Prior findings also suggest that negative attitudes towards cancer screening, including fear of pain or diagnosis, disbelief in the efficacy of screening tests, or generalized distrust of others may be more predominant among racial and ethnic minorities and thus account for their lower rates of cancer screening.[16,17,21,23-25,29,30] It may be that such factors, which are closely associated with race, actually drive the race differences in cancer screening, but whether this is the case has not been well examined. An additional limitation of prior research on various minority populations which has examined the effects of attitudes and beliefs about cancer screening is that it has often included only limited sample sizes from single geographic areas. Further, few studies have specifically examined either the effects of having cancer-related symptoms (or not), or the type of setting in which cancer screening is provided, on individuals' willingness to have screening. Presentation of symptoms and screening setting both have the potential to impact the decision to seek testing,[31-33] so attention to these issues is important for a full understanding of patient attitudes about cancer screening.

Although it is important to identify barriers for specific subsets of the population for different types of cancer screening, there is also value in understanding general attitudes of patients about cancer screening in general in order to highlight common barriers for future research and interventions. While 2 recent papers by our research group addressed this latter issue and reported that blacks and Hispanics, after adjusting only for demographic factors, were either as or more likely than whites to self-report willingness to have cancer screenings, and perceived a higher risk of "not getting a thorough cancer screening" than did whites, neither of those analyses delved into underlying reasons for those observed differences.[34,35] Therefore, the purpose of this current analysis is to examine general willingness to engage in cancer screening, in the context of varied symptoms and screening settings, and among a diverse cohort from multiple geographic areas, with an emphasis on assessing the influence of race/ethnicity, relative to sociodemographic factors, to both positive and negative attitudes and beliefs about cancer screening on this willingness to utilize screening in general. We hypothesized that the association between race and willingness to receive cancer screening would be attenuated after taking into account a wide variety of attitudes, barriers, and facilitators to cancer screening, and after accounting for positive attitudes for screening.

Further, we anticipated that in the setting of both the greatest urgency (eg, the presence of symptoms) and most accessible and comfortable screening location (eg, one's doctor's office), the effects of barriers and facilitators to screening would be minimized. Our work was guided by the health decision model (HDM),[36] which is an expanded version of the patient-focused health belief model[37] and includes factors beyond the patient's own attitudes and beliefs, which might influence health decisions, such as sociodemographic factors, experiences with the health care system, and knowledge. In the HDM, there is not a particular causal ordering of factors influencing health decisions; rather, each of the domains of health beliefs, patient preferences, experience, and knowledge influence one another, and are also affected by social interaction and sociodemographic factors. Thus, in this study, we used the HDM as a guide for thinking about the various sociodemographic factors, barriers, facilitators hypothesized to affect willingness to have cancer screening, but not for positing causal relationships among the elements.

METHODS

Study Sample and Procedures

To ensure a wide geographic, racial/ethnic representation which included substantial representation of whites, African Americans, and Hispanics, we contacted a random sample of residents of 3 cities—San Juan, Puerto Rico; Baltimore, Maryland; and New York, New York—from September to December 2003. We conducted random-digit-dial telephone interviews to noninstitutionalized adults residing in telephone-equipped homes. A total of 1148 adult African Americans, Puerto-Rican Hispanics, and whites responded, with response rates of 58%, 51%, and 45%, respectively, and an overall completion rate of 82.6%. The final study sample consisted of 356 African Americans, 313 Puerto Rican Americans, and 493 non-Hispanic whites. Since we did not collect identifying information about respondents, our institutional review board (IRB) determined this study to be "exempt" from full IRB review.

MEASURES

Questionnaire

We administered a questionnaire to all study participants, which contained all measures used in this analysis, and which has been fully described (as regards development, administration, and data analysis decisions) in our prior publications based upon this survey.[34,35,38] The questionnaire was administered either in English or Spanish, at the preference of the respondent.

Dependent Variables

Willingness to have cancer screening under spe-

cific conditions. We asked questions about individuals' self-reported willingness to participate in cancer screening exams under 4 different conditions. These questions assessed the respondent's self-reported likelihood to have a cancer screening: (1) by their own *physician* when there are *no* symptoms ("doctor/no symptoms;" the question read, "How likely are you to go for a regular annual cancer screening exam given by your doctor, if you have NO symptoms?"); (2) by their own *physician* when there *are* symptoms ("doctor/symptoms;" the question was: "If your own doctor told you that you have some symptoms and needed a cancer screening exam, how likely are you to go and have *that* cancer screening exam?"); (3) when it is a free cancer screening exam in the *community* when there are *no* symptoms ("community/no symptoms:" the question read "Some group in your community, such as a school, church, or the Lions' Club, offers you the opportunity to have a free cancer screening exam. How likely are you to participate at a community-level *free* cancer screening exam if you *have no symptoms?"*); and (4) when it is a free cancer screening exam in the *community* when there *are* symptoms ("community/symptoms;" the question was: "Some group in your community, such as a school, church, or the Lions' Club, offers you the opportunity to have a free cancer screening exam. How likely are you to participate at a community level *free* cancer screening exam if you *have symptoms*?"). Responses for all questions were on a 5-point scale with 1 signifying "very unlikely" and 5 signifying "very likely." The respondent was not given

Table 1. Characteristics of Study Sample and Association of Sociodemographics With Dependent Variables

				Mean of Each Dependent Variable by Sociodemographic Characteristics			
	N	%	95% CI	MD/ No Symptoms	Community/ No symptoms	Community/ Symptoms	MD/ Symptoms
				P = .02	*P* = .61	*P* = 0.04	*P* = .35
Male	401	45.5	41.2-49.8	3.59	3.28	4.06	4.81
Female	747	54.5	50.2-58.8	3.89	3.21	3.78	4.87
Race/ethnicity				*P* = .20	*P* < .0001	*P* < .0001	*P* = .52
Puerto Rican Hispanic	311	22.4	19.1-25.7	3.85	3.70	4.32	4.86
Black	355	31.1	27.3-35.0	3.87	3.69	4.23	4.89
White	482	46.5	42.5-50.4	3.63	2.73	3.50	4.81
Age				*P* < .0001	*P* = .09	*P* < .0001	*P* = .43
18-39	216	22.9	19.9-27.4	3.27	3.45	4.35	4.79
40-59	679	53.4	49.2-57.7	3.78	3.27	3.98	4.88
≥60	253	23.7	19.9-27.5	4.17	2.97	3.32	4.82
Income				*P* = .41	*P* < .0001	*P* < .0001	*P* = .07
<$20 000	327	24.2	20.5-27.8	3.74	3.62	4.37	4.83
$20 000-$34 999	246	24.5	20.6-28.4	3.59	3.59	4.15	4.82
$35 000-$49 999	155	16.0	12.7-19.3	3.95	3.32	3.68	4.81
$50 000-$74 999	149	16.7	13.3-20.2	3.62	2.98	3.81	4.95
≥$75 000	146	18.6	15.0-22.2	3.88	2.56	3.40	4.83
Education				*P* = .02	*P* < .0001	*P* = .01	*P* = .08
High school or less	182	12.1	9.3-14.8	3.67	3.72	4.20	4.83
Tech school	312	27.1	23.3-30.9	3.69	3.44	4.04	4.77
Some college	241	22.8	19.1-26.5	3.54	3.25	3.98	4.86
College grad	252	23.9	20.3-26.6	3.87	3.20	3.89	4.89
Postgraduate	153	14.1	11.1-17.1	4.12	2.55	3.35	4.95
Employed				*P* = .26	*P* = .78	*P* = .64	*P* = .28
Yes	666	59.6	55.4-63.7	3.70	3.22	3.94	4.87
No	479	40.4	36.3-44.6	3.84	3.26	3.88	4.81
Health status				*P* = .83	*P* = .18	*P* = .35	*P* = .60
Excellent	216	19.4	16.0-22.8	3.76	3.25	4.01	4.91
Very good	354	33.3	29.3-37.2	3.69	3.11	3.73	4.81
Good	333	27.8	24.0-31.5	3.76	3.20	3.97	4.84
Fair	194	15.9	12.7-19.2	3.90	3.59	4.09	4.88
Poor	49	3.6	1.2-6.1	3.83	3.10	3.79	4.67

a definition of what "by your doctor" referred to, so responses are based on their subjective interpretation of each item.

Independent Variables

Barriers to and facilitators of cancer screening. The questionnaire also included items to assess the impact of attitudes and beliefs about potential *barriers* to screening, including fear of: (1) getting AIDS; (2) being a "guinea pig;" (3) test results not being private or confidential; (4) how the disease would upset one's family; (5) hearing one has cancer; (6) feeling that one is unlikely to get cancer; (7) lack of trust in medical professionals; (8) fear that the test might be painful; and (9) fear of being embarrassed in the cancer screening exam. In addition, the questionnaire asked about factors that might *facilitate* an individual's participation in cancer screening, including: (1) the belief that early detection might save one's life; (2) having close friends or a relative encouraging participation in cancer screening; (3) having close friends or a relative participating in cancer screening; (4) having a close friend or relative who has had cancer; (5) encouragement of one's physician to be screened; (6) encouragement of one's dentist to be screened; (7) having one's insurance company paying for the screening exam; and (8) having a nearby location for the screening. A 5-point Likert scale was used for responses to all questions, ranging from "not at all" to "totally." These were analyzed as single items.

Sociodemographic factors. Respondents were asked to self-report their sex, race, ethnicity (Puerto Rican Hispanic or not), year of birth, income (in $5000 increments), highest education attained, whether they were currently employed, and health status (response categories: excellent, very good, good, fair, poor).

Covariates

General attitudes. We assessed respondents' general attitudes towards cancer screening exams, asking how effective the respondent believed cancer screening exams are in detecting cancer (higher scores indicate stronger beliefs in effectiveness). In addition, to understand if willingness to participate in cancer screening might be associated with a person's general trust in people, we included the Trust in People scale.[39] This scale includes the following questions: Generally speaking, would you say that: (1) most people can be trusted or that you can't be too careful in dealing with people?; (2) most of the time, people try to be helpful, or that they are mostly just looking out for themselves?; (3) most people would try to take advantage of you if they got the chance or would they try to be fair? The scale is the sum of these 3 items (range, 0-3); a higher score indicates greater trust in people.

Statistical Analysis

Initially, we examined descriptive statistics for all variables, including distributions of responses to each of the dependent variables by race/ethnicity. To explore

Table 2. Zero-Order Correlations Between Each Barrier and Facilitator and the Dependent Variables

	MD/No Symptoms		Community/ No Symptoms		Community/ Symptoms		MD/Symptoms	
	R	p	R	p	R	p	R	p
How effective screening	-0.18	<.0001	-0.01	.83	-0.06	.22	-0.16	.003
Trust	0.11	.01	-0.20	<.0001	-0.22	<.0001	0.11	.03
Barriers								
AIDS	-0.02	.431	0.11	.012	0.11	.014	0.01	.867
Guinea pig	-0.03	.428	0.02	.614	0.08	.076	0.08	.055
Privacy	-0.01	.521	-0.03	.481	0.03	.427	0.05	.296
Upsetting family	-0.08	.730	0.12	.003	0.10	.017	0.05	.244
Fear of cancer	-0.03	.317	0.06	.139	0.10	.019	-0.01	.896
Unlikely to get cancer	-0.10	.639	0.03	.396	0.11	.007	0.04	.326
Lack trust in medicine	0.03	.380	0.11	.008	0.07	.100	-0.04	.282
Fear of pain	-0.07	.909	0.08	.052	0.08	.067	0.01	.904
Embarrassment	-0.16	.002	0.01	.773	0.06	.145	-0.06	.193
Facilitators								
Saves lives	0.26	<.0001	0.17	.000	0.12	.017	0.19	.024
Friends/family encouraged	0.12	.014	0.13	.005	0.12	.008	0.11	.056
Friends participate	0.13	.014	0.14	.003	0.11	.016	0.15	.015
Friends/relative w/cancer	0.12	.000	0.10	.030	0.14	.003	0.20	.004
MD encourages	0.25	<.0001	0.01	.760	0.00	.995	0.21	.008
Dentist encourages	0.19	<.0001	0.04	.346	0.01	.899	0.16	.020
Insurance	0.16	.013	0.14	.002	0.09	.057	0.17	.012
Screening nearby	0.12	.002	0.23	<.0001	0.20	<.0001	0.13	.042

whether the dependent variables regarding willingness to have cancer screening would function better as a scale or several scales, we conducted exploratory factor analyses but found that the factors did not lead to scales with acceptable internal consistency reliability. Thus, we retained 4 separate items for the outcome variables. Next, we examined bivariate associations between each independent variable and the dependent variables, examining dependent variable means by each category of the independent sociodemographic variables and 0-order correlations between each barrier or facilitator and the dependent variables. Then, we computed multiple linear regression models to examine the effects of race/ethnicity and other sociodemographic factors, after adjusting for the barriers and facilitators to willingness to have cancer screening, under the four conditions.

RESULTS

Sociodemographic Characteristics of the Sample

The mean age of respondents was 44.9 years (not shown); 54.5% of the sample were female (Table 1). Puerto Rican Hispanics comprised 22.4% (n = 311) of the sample, while 31.1% were non-Hispanic African American (n = 355) and 46.5% were non-Hispanic white (n = 482). Regarding income, 24.2% earned less then $20 000; 24.5% earned between $20 000 and 34 999; 16.0% earned between $35 000 and $49 999; 16.7% earned between $50 000 and $74 999; and 18.6% earned more than $75 000. More than half of the sample (59.6%) was employed, 12.1% had a high school education or less, and 52.7% indicated a health status of excellent or very good.

Ranking of the 4 Scenarios by Willingness to Have Cancer Screenings by Race/Ethnicity

To examine the overall likelihood of respondents indicating they were "very likely" to obtain screening within the 4 scenarios (screening by one's own physician with and without symptoms; and screening in the community with and without symptoms), we found that the 4 scenarios were ranked in the same order by African Americans, Puerto Rican Hispanics, and whites: (1) "own MD, with symptoms" (with 92%-93% responding "very likely" across the racial/ethic groups); (2) "community screening, with symptoms" (47%-71%); (3) "own MD, no symptoms" (44%-50%), and (4) "community event, no symptoms" (23%-45%) (results not shown). With the exception of the first scenario, in which greater than 90% of each of the 3 racial/ethnic groups responded "very likely," Puerto Rican Hispanics were the most willing to have a cancer screening under each of the 3 other scenarios ($p \leq .0001$).

Factors Associated With Willingness to Have Cancer Screening

Screening by one's own physician given no symptoms. Bivariate analyses indicated that female sex, older age, and more education were associated with more willingness for cancer screening by one's own physician when there are no symptoms (Table 1). Beliefs in one's likelihood of getting cancer, fear of pain or embarrassment upon screening, the perception that screening saves lives, friends/family encouraged one to get screened, also participate in screening, and have had cancer, encouragement from one's physician or dentist, having insurance and screening nearby were all significantly associated with willingness to have screening (Table 2).

When we then adjusted for the effects of potential barriers and facilitators of screening by one's own physician given no symptoms, we found that the barrier of fear of getting cancer was associated with a greater likelihood of having screening, while beliefs about cancer screening's effectiveness, perceptions of likelihood of getting cancer, less concern about being embarrassed by the screening exam, and perceiving that screening saves lives were associated with the likelihood of having screening (Table 3). Although education was no longer significant in the multivariate model, we found that those with income levels between $50 000 and $75 000 were less likely to be willing to have cancer screening by one's own physician, compared to those of the highest income level, as were younger persons and males, while African Americans were more willing to have screening than whites.

Free community screening given no symptoms. Bivariate results (Table 1) indicated that Puerto Rican Hispanics and African Americans were more willing to obtain cancer screening in the community, given no symptoms, than were whites, as were those with lower incomes. In addition, those with the lower levels of education were more likely to express willingness to be screened in a community setting given no symptoms than those with more education. Fear of getting AIDS, cancer, lack of trust in the medical establishment, fear of pain, as well as the perception that screening saves lives, friends/family encouraged one to get screened, also participate in screening, and have had cancer, having insurance and screening nearby were all significantly associated with willingness to have screening (Table 2).

In the multivariate model, only 1 barrier—embarrassment—remained significant after adjustment for other factors (greater embarrassment was associated with less likelihood of screening). Two facilitators were significantly associated with screening willingness—physician encouragement, which was associated with less likelihood to get screening, and having screening nearby, which was associated with a greater willingness to be screened. Being Puerto Rican Hispanic or African American, having lower income and less education were

each associated with greater willingness for community cancer screening with no symptoms.

Free community screening given symptoms. In bivariate analyses, almost all sociodemographic factors were associated with screening willingness in this context: being male, younger, being Puerto Rican Hispanic or African American, having lower income, and less education were associated with greater likelihood of indicating willingness for community screening given symptoms. Fear of getting AIDS or cancer, or having pain, as well as the perception that screening saves lives, friends/family encouraged one to get screened, also par-

Table 3. Bivariate and Multivariate Analyses for Each Dependent Screening Variable

	Screened by MD/ No Symptoms Q8				Screened by Community Group/No Symptoms q9			
	Bivariate Associations		Multivariate Associations		Bivariate Associations		Multivariate Associations	
	Coeff	*P*	Coeff	*P*	Coeff	*P*	Coeff	*P*
Male	-0.3	.019	-0.32	.028	0.07	.606	0.28	.079
Female	0		0		0		0	
Hispanic	0.21	.174	0.23	.221	0.97	<.0001	0.69	.001
Black, non-Hispanic	0.24	.109	0.36	.028	0.96		0.53	.009
White, non-Hispanic	0		0		0		0	
18-39	-0.9	<.0001	-0.95	<.0001	0.47	0.027	0.44	.108
40-59	-0.39	.007	-0.32	.048	0.3	1.113	0.30	.188
≥60	0		0		0		0	
<$20 000	-0.15	.436	0.18	.454	1.06	<.0001	0.86	.003
$20 000-$34 999	-0.29	.153	-0.01	.958	1.03	<.0001	0.84	.001
$35 000-$49 999	0.07	.770	0.28	.223	0.76	0.003	0.61	.018
$50 000-$74 999	-0.26	.261	-0.50	.024	0.42	.089	0.49	.043
≥$75 000	0		0		0		0	
High school or less	-0.45	.021	-0.21	.412	1.17	<.0001	0.64	.041
Technical school	-0.43	.015	-0.44	.066	0.88	<.0001	0.37	.172
Some college	-0.59	.003	-0.27	.256	0.7	.002	0.40	.114
College graduate	-0.25	.151	0.02	.913	0.65	.003	0.63	.011
Postgraduate	0		0		0		0	
Employed	-0.14	.258	0.08	.637	-0.04	.776	-0.01	.952
Not Employed	0		0		0		0	
Health status	0.04	.427	-0.05	.435	0.07	0.221	-0.11	.135
How effective screening	-0.31	<.0001	-0.20	.034	-0.02	.826	-0.05	.528
Trust in People	0.15	.015	0.11	.121	-0.29	<.0001	-0.07	.377
Barriers								
AIDS	-0.02	.624	0.05	.361	0.12	.012	0.06	.278
Guinea pig	-0.02	.535	-0.06	.185	0.02	.614	-0.03	.588
Privacy	-0.01	.884	0.08	.116	-0.03	.481	-0.11	.055
Upsetting family	-0.07	.081	-0.10	.134	0.13	.003	0.09	.172
Fear of cancer	-0.03	.456	0.16	.007	0.07	.139	-0.01	.878
Unlikely to get cancer	-0.1	.032	-0.15	.011	0.04	.396	-0.11	.078
Lack trust in medicine	0.03	.538	0.10	.084	0.13	.008	0.06	.362
Fear of pain	-0.08	.091	-0.05	.458	0.1	.052	0.10	.168
Embarrassment	-0.2	.000	-0.18	.029	0.02	.773	-0.15	.032
Facilitators								
Saves lives	0.36	<.0001	0.23	.029	0.26	.000	0.08	.321
Friends/family encourage	0.13	.012	0.08	.326	0.15	.005	-0.03	.709
Friends participate	0.13	.009	0	.946	0.16	.003	0.11	.238
Friends/family w/cancer	0.14	.008	0.03	.607	0.12	.030	-0.01	.930
MD encourages	0.31	<.0001	0.11	.271	0.02	.760	-0.22	.031
Dentist encourages	0.2	<.0001	-0.03	.638	0.05	.346	0.10	.286
Insurance	0.17	.001	0.03	.658	0.16	.002	-0.01	.927
Screening nearby	0.13	.010	-0.06	.283	0.27	<.0001	0.19	.009

ticipate in screening, and have had cancer, having insurance and screening nearby were all significantly associated with willingness to have screening (Table 2).

In the final model, no barriers were significant, but beliefs in the effectiveness of screening were significantly associated with the outcome, and having screening nearby was associated with greater willingness for screening. Male sex, being Puerto Rican Hispanic, being of younger age, and having lower income were also all positively associated with willingness for community screening given symptoms.

Screening by one's own doctor given symptoms. No sociodemographic variables were associated with the likelihood of getting screening by one's own physician, given symptoms, in bivariate analyses (Table 1). Fear of begin a guinea pig, the perception that screening saves lives, having family or friends who have had cancer, having a physician or dentist encourage one to have screening, and having insurance were all significantly associated with willingness to have screening (Table 2).

In the full model (Table 3), no barriers or facilitators were significantly associated with screening willingness. Only being employed and having greater trust in people were significantly associated with increased willingness to participate in cancer screening by one's own physician, given symptoms; there were no significant effects of race/ethnicity.

Screened by Community Group/Symptoms q9a				Screened by MD/ Symptoms q10			
Bivariate Associations		Multivariate Associations		Bivariate Associations		Multivariate Associations	
Coeff	*P*	Coeff	*P*	Coeff	*P*	Coeff	*P*
0.27	.043	0.39	.014	-0.06	.354	-0.03	.707
-	-	0		0		0	
0.82	<.0001	0.75	<.0001	0.04	.526	0.05	.446
0.73	<.0001	0.35	.079	0.07	.256	0.07	.404
-	-	0		0		0	
1.03	<.0001	1.24	<.0001	-0.03	.770	-0.04	.771
0.66	.000	0.94	<.0001	0.06	.484	0.03	.747
-	-	0		0	-	0	
0.97	<.0001	0.61	.015	0	.970	0.11	.408
0.75	.001	0.53	.027	-0.01	.888	0.07	.486
0.28	.291	0.07	.779	-0.02	.887	0.05	.659
0.41	.120	0.19	.486	0.12	.158	0.03	.698
-	-	0		0	-	0	
0.84	.000	-0.04	.900	-0.12	.089	0.00	.973
0.68	.003	0.11	.696	-0.18	.020	-0.12	.292
0.63	.010	0.17	.507	-0.10	.127	-0.04	.690
0.54	.024	0.35	.151	-0.06	.330	0.03	.681
-	-	0		0	-	0	
0.06	.642	-0.21	.211	0.07	.283	0.21	.006
-	-	0		0	-	0	
0.03	.551	0.07	.294	-0.02	.498	0.00	.963
-0.1	.222	-0.26	.002	-0.12	.003	-0.07	.191
-0.31	<.0001	-0.12	.100	0.07	.026	0.09	.045
-	-	-	-		-		
0.11	.014	0.07	.227	0	.867	-0.02	.262
0.08	.076	0.03	.556	0.03	.055	0.02	.166
0.03	.427	-0.07	.237	0.02	.296	0.04	.059
0.1	.017	-0.06	.465	0.02	.244	0.01	.734
0.1	.019	0.04	.546	0	.896	0.00	.918
0.13	.007	0.05	.402	0.02	.326	0.04	.101
0.08	.100	-0.04	.585	-0.02	.282	-0.03	.342
0.09	.067	0.06	.393	0	.904	-0.02	.389
0.08	.145	-0.04	.556	-0.03	.193	0.01	.707
-	-				-		
0.18	.017	0.03	.717	0.11	.024	0.05	.414
0.14	.008	-0.02	.715	0.05	.056	-0.01	.831
0.12	.016	-0.01	.915	0.07	.015	-0.01	.653
0.16	.003	0.05	.498	0.10	.004	0.07	.069
0	.995	-0.14	.110	0.12	.008	0.03	.577
0.01	.899	0	.996	0.07	.020	0.02	.339
0.1	.057	-0.02	.744	0.08	.012	0.00	.886
0.22	<.0001	0.18	.004	0.06	.042	0.00	.902

SUMMARY OF RESULTS

Across the multivariate models for all dependent variables, health status was never significantly associated with screening willingness, and being employed was only associated with willingness for screening by one's physician when one has symptoms. Respondents' ratings of the effectiveness of screening were associated with likelihood of screening in 3 of the 4 scenarios. Trust in people was associated with screening willingness (in the context of having symptoms, and screening by one's physician such that more trust was associated with greater willingness to be screened). In bivariate results, Puerto Rican Hispanics and/or African Americans were significantly more willing than whites to have all types of screening, except for the scenarios in which screening would be done by one's own physician with or without symptoms (there were no race/ethnic differences and very few sociodemographic differences overall in endorsement of this outcome, with 92% to 93% of each racial/ethnic group indicating "very likely" for this specific scenario). Thus, some effects of race/ethnicity persisted across almost all

multivariate models (except for screening by a physician when symptoms existed), such that the effect remained, although it was slightly attenuated, even after adjusting for other sociodemographic factors, barriers and facilitators of screening.

DISCUSSION

We examined the effects of race/ethnicity and other sociodemographic characteristics on willingness for cancer screening, after accounting for the effects of attitudes about potential barriers and facilitators of cancer screening, drawing on data from a diverse sample. We considered these dynamics in the context in which screening would be done (community vs by one's own physician) and in the context of whether or not the respondent had cancer-related symptoms.

On the bivariate level, numerous sociodemographic factors were associated with willingness to have screening, with the exception of physician-administered screening when there were symptoms, where no sociodemographic variables were significant. Notably, the effects of the sociodemographic factors (especially race/ethnicity and income) consistently persisted in multivariate models. This suggests that effects of these characteristics are not attenuated by the inclusion of attitudes and beliefs, and points to the probable independent impact of these characteristics on screening willingness. The relative size of the effects we observed also supports this notion, as the size of the effects for sociodemographics were generally greater than those for the attitudinal/belief variables.

However, across the models, several barriers and facilitators of screening were associated with screening willingness, again with the exception of physician-administered screening with symptoms present. Thus, we conclude that sociodemographic factors are associated with willingness to have cancer screening in almost all cases, but that perceived barriers and facilitators also matter sometimes, as well.

Almost no variables in our multivariate models were significantly associated with willingness to have screening done by one's own physician given the presence of cancer-related symptoms. This suggests that the urgency or concern associated with such a screening, to be conducted in the relative privacy and familiarity of one's own doctor's office, overshadows any of the attitudinal or sociodemographic dimensions we measured. Thus, in situations where screening seems less "discretionary," neither sociodemographic factors nor attitudes contributed strongly to willingness for screening. Similarly, for community screenings in the presence of symptoms, almost no attitudes were significant, although sociodemographics were.

Being Puerto Rican Hispanic or African American was a fairly consistent predictor of willingness to have cancer screening. Thus, race/ethnicity may be a factor that needs to be considered in the case of public health outreach for cancer screening, but it seems less important in the context of individual clinicians recommending cancer screening within their own setting. These findings also support the idea that making free cancer screening available in the community will help to attract more African Americans and Hispanics.

The strengths of this study included the focus on an ethnically and racially diverse sample from multiple geographic areas, the availability of data on both sociodemographic and attitudinal factors and the inclusion of multiple questions about cancer screening in a variety of contexts. While it can be argued that this study was limited by its reliance on questions about potential willingness to seek cancer screening in general (vs actual receipt of screening), prior health behavior research has shown that intentions for health behavior are important predictors of actual health behavior.[40] There is also value in understanding individuals' beliefs about cancer screening in general, different from their thoughts about specific cancers and screening tests for them. We were unable to account for the effects of having a primary care physician, compared to not, which may impact willingness to seek screening, although our questions did ask about willingness to have cancer screening by one's own physician. Similarly, we did not have data on the proportion of patients having their own primary care physician available for this analysis, which is a limitation of the study. Our questionnaire also did not ask about other potential factors associated with screening, such as history of screening and family history of cancer. The general Trust in People scale may not translate into trust in the medical community, yet we felt it important to account for individuals' general levels of trust, which would likely affect their trust in the medical establishment as well. As previously noted, our questionnaire addressed cancer screening in general, though there is some evidence that willingness to screen may vary with the type of cancer. However, information of general attitudes towards cancer screening may be helpful in designing future interventions and campaigns to improve overall screening rates.

Our results document the important effects of sociodemographic factors on willingness to have cancer screening and echo others' findings that knowledge and attitudes about 1 particular cancer screening—mammography—did not independently predict its use.[41] As others have noted, more needs to be known about cancer screening practices among Hispanics in the United States.[4] The results of this investigation clearly show that Puerto Rican Hispanic ethnicity is an important predictor of willingness to have community based screening. This study contributes new information to the literature indicating that the relationship between race/ethnicity and willingness to be screened are not attenuated by attitudes about potential barriers and facilitators to screening, and also that screening site may influence individuals' willingness to be screened.

In summary, our results indicate that willingness to seek cancer screening is influenced by sociodemographic characteristics, over and above attitudes about screening, and yet, in some contexts, these latter factors should also be considered in promoting screenings. Since most sociodemographic characteristics are not easily mutable, their potential impact on cancer screening availability and awareness campaigns must be recognized, so as to target such campaigns to the populations which can most benefit from the needed cancer screening. In addition, it appears that both the location of screening and the potential urgency of screening influence the relative importance of each type of factor, suggesting that future efforts to increase screening should consider location and emphasize urgency, where appropriate.

ACKNOWLEDGMENT

We thank Karen Clements, PhD, for data analysis.

REFERENCES

1. Swan J, Breen N, Coates R, et al. Progress in cancer-screening practices in the United States: results from the 2000 National Health Interview Survey. *Cancer*. 2003;97(6):1528-1540.

2. Zambrana R, Breen N, Fox S, et al. Use of cancer-screening practices by Hispanic women: analyses by group. *Prev Med*. 1999;29:466-477.

3. American Cancer Society. *Cancer Prevention & Early Detection Facts & Figures 2009*. Atlanta, GA: American Cancer Society; 2009.

4. Selvin E, Brett K. Breast and cervical cancer screening: Sociodemographic predictors among White, Black and Hispanic women. *Am J Public Health*. 2003;93:618-623.

5. Kerlikowske K, Grady D, Rubin SM, et al. Efficacy of screening mammography. A meta-analysis. *JAMA*. 1995;273(2):149-154.

6. Sasieni P, Adams J. Effect of screening on cervical cancer mortality in England and Wales: analysis of trends with an age period cohort model. *BMJ*. 1999;318(7193):1244-1245.

7. Mandel JS, Bond JH, Church TR, et al. Reducing mortality from colorectal cancer by screening for fecal occult blood. Minnesota Colon Cancer Control Study. *N Engl J Med*. 1993;328(19):1365-1371.

8. Jemal A, Siegel R, Ward E, et al. Cancer Statistics, 2009. *CA Cancer J Clin*. May 27 2009.

9. Morris GJ, Mitchell EP. Higher incidence of aggressive breast cancers in African-American women: a review. *J Natl Med Assoc*. 2008;100(6):698-702.

10. Jones BA, Liu WL, Araujo AB, et al. Explaining the race difference in prostate cancer stage at diagnosis. *Cancer Epidemiol Biomarkers Prev*. 2008;17(10):2825-2834.

11. Kothari A, Fentiman IS. 22. Diagnostic delays in breast cancer and impact on survival. *Int J Clin Pract*. 2003;57(3):200-203.

12. Coughlin SS, Leadbetter S, Richards T, et al. Contextual analysis of breast and cervical cancer screening and factors associated with health care access among United States women, 2002. *Soc Sci Med*. 2008;66(2):260-275.

13. Guerra CE, Schwartz JS, Armstrong K, et al. Barriers of and facilitators to physician recommendation of colorectal cancer screening. *J Gen Intern Med*. 2007;22(12):1681-1688.

14. Shokar NK, Carlson CA, Weller SC. Factors associated with racial/ethnic differences in colorectal cancer screening. *J Am Board Fam Med*. 2008;21(5):414-426.

15. Andrasik MP, Rose R, Pereira D, et al. Barriers to cervical cancer screening among low-income HIV-positive African American women. *J Health Care Poor Underserved*. 2008;19(3):912-925.

16. Odedina FT, Campbell ES, LaRose-Pierre M, Scrivens J, Hill A. Personal factors affecting African-American men's prostate cancer screening behavior. *J Natl Med Assoc*. 2008;100(6):724-733.

17. Green AR, Peters-Lewis A, Percac-Lima S, et al. Barriers to screening colonoscopy for low-income Latino and white patients in an urban community health center. *J Gen Intern Med*. 2008;23(6):834-840.

18. Mishra SI, Bastani R, Huang D, Luce PH, Baquet CR. Mammography screening and Pacific Islanders: role of cultural and psychosocial factors. *J Cancer Educ*. 2007;22(1):32-36.

19. Gany FM, Herrera AP, Avallone M, Changrani J. Attitudes, knowledge, and health-seeking behaviors of five immigrant minority communities in the prevention and screening of cancer: a focus group approach. *Ethn Health*. 2006;11(1):19-39.

20. McGarvey EL, Clavet GJ, Johnson JB, 2nd, Butler A, Cook KO, Pennino B. Cancer screening practices and attitudes: comparison of low-income women in three ethnic groups. *Ethn Health*. 2003;8(1):71-82.

21. Austin LT, Ahmad F, McNally MJ, Stewart DE. Breast and cervical cancer screening in Hispanic women: a literature review using the health belief model. *Wom Health Issues*. 2002;12(3):122-128.

22. Lubetkin EI, Santana A, Tso A, Jia H. Predictors of cancer screening among low-income primary care patients. *J Health Care Poor Underserved*. 2008;19(1):135-148.

23. Consedine NS, Magai C, Krivoshekova YS, Ryzewicz L, Neugut AI. Fear, anxiety, worry, and breast cancer screening behavior: a critical review. *Cancer Epidemiol Biomarkers Prev*. 2004;13(4):501-510.

24. Consedine NS, Morgenstern AH, Kudadjie-Gyamfi E, Magai C, Neugut AI. Prostate cancer screening behavior in men from seven ethnic groups: the fear factor. *Cancer Epidemiol Biomarkers Prev*. 2006;15(2):228-237.

25. Otero-Sabogal R, Stewart S, Sabogal F, Brown B, Perez-Stable E. Access and attitudinal factors related to breast and cervical cancer rescreening. Why are Latinas still underscreened? *Health Educ Behav*. 2003;30(3):337-359.

26. Abraido-Lanza A, Chao M, Gammon M. Breast and cervical cancer screening among Latinas and Non-Latina whites. *Am J Public Health*. 2004;94:1393-1398.

27. Lillie-Blanton M, Laveist T. Race/ethnicity, the social environment, and health. *Soc Sci Med*. 1996;43(1):83-91.

28. Mayberry RM, Mili F, Ofili E. Racial and ethnic differences in access to medical care. *Medical Care Research and Review*. 2000;57(suppl 1):108-145.

29. Luquis R, Villanueva I. Knowledge, attitudes, and perceptions about breast cancer and breast cancer-screening among Hispanic women residing in South Central Pennsylvania. *J Community Health*. 2006;31(1):25-42.

30. Hubbell F, Chavez L, Mishra S, Valdez R. Differing beliefs about breast cancer among Latinas and Anglo women. *West J Med*. 1996;164:405-409.

31. Romero FR, Romero KR, Brenny FT, Pilati R, Kulysz D, de Oliveira Junior FC. Reasons why patients reject digital rectal examination when screening for prostate cancer. *Arch Esp Urol*. 2008;61(6):759-765.

32. Lasser KE, Ayanian JZ, Fletcher RH, Good MJ. Barriers to colorectal cancer screening in community health centers: a qualitative study. *BMC Fam Pract*. 2008;9:15.

33. Ackerson K, Gretebeck K. Factors influencing cancer screening practices of underserved women. *J Am Acad Nurse Pract*. 2007;19(11):591-601.

34. Katz RV, Wang MQ, Green BL, et al. Participation in biomedical research studies and cancer screenings: perceptions of risks to minorities compared with whites. *Cancer Control*. 2008;15(4):344-351.

35. Katz RV, Claudio C, Kressin NR, Green BL, Wang MQ, Russell SL. Willingness to participate in cancer screenings: blacks vs whites vs Puerto Rican Hispanics. *Cancer Control*. 2008;15(4):334-343.

36. Rosenstock I. Why people use health services. *Milbank Memorial Fund Q*. 1966;44(94).

37. Eraker SA, Kirscht JP, Becker MH. Understanding and improving patient compliance. *Ann Intern Med*. 1984;100:258-268.

38. Claudio C, Katz R, Green B, Kressin N, Wang MQ, Russell S. Cancer-Screening Participation: Comparative Willingness of San Juan Puerto Ricans vs. New York City Puerto Ricans. *J Natl Med Assoc*. in press.

39. Survey Research Center. *1964 Election Study*. Ann Arbor, MI: Inter-University Consortium for Political Research, University of Michigan; 1969.

40. Ajzen I, Fishbein M. *Understanding attitudes and predicting social behavior*. Englewood Cliffs, NJ: Prentice-Hall; 1980.

41. Hubbell F, Mishra S, Chavez L, Valdez R. The influence of knowledge and attitudes about breast cancer on mammography use among Latinas and Anglo women. *J Gen Intern Med*. 1997;12:505-508. ■

Understanding Contributors to Racial Disparities in Blood Pressure Control

Nancy R. Kressin, PhD; Michelle B. Orner, MPH; Meredith Manze, MPH;
Mark E. Glickman, PhD; Dan Berlowitz, MD, MPH

Background—Racial disparities in blood pressure (BP) control are well documented but poorly understood; prior studies have only included a limited range of potential explanatory factors. We examined a comprehensive set of putative factors related to blood pressure control, including patient clinical and sociodemographic characteristics, beliefs about BP and BP medications, medication adherence, and experiences of discrimination, to determine if the impact of race on BP control remains after accounting for such factors.

Methods and Results—We recruited 806 white and black patients with hypertension from an urban safety-net hospital. From a questionnaire administered to patients after their clinic visits, electronic medical record and BP data, we assessed an array of patient factors. We then examined the association of patient factors with BP control by modeling it as a function of the covariates using random-effects logistic regression. Blacks indicated worse medication adherence, more discrimination, and more concerns about high BP and BP medications, compared with whites. After accounting for all factors, race was no longer a significant predictor of BP control.

Conclusions—Results suggest that equalizing patients' health beliefs, medication adherence, and experiences with care could ameliorate disparities in BP control. Additional attention must focus on the factors associated with race to identify, and ultimately intervene on, the causes of racial disparities in BP outcomes. **(*Circ Cardiovasc Qual Outcomes*. 2010; 3:173-180.)**

Key Words: blood pressure ■ hypertension ■ race ■ disparities

Hypertension, which affects more than 73 million Americans, is a major risk factor for cardiovascular, cerebrovascular, and renal disease.[1] It is more frequent among African Americans[2] and accounts for a significant portion of racial differences in mortality, through excess cardiovascular morbidity.[3] Many patients with hypertension have poorly controlled blood pressure (BP), and African Americans are disproportionately represented among this group,[4] even after controlling for comorbidities such as diabetes and renal disease.[4–7]

The reason for this racial disparity is not well understood. Many prior studies of BP control have only examined a narrow range of potential etiologic factors—usually clinical characteristics and sometimes including selected sociodemographic factors.[8,9] Recently, authors have suggested that patient self-management attitudes and behaviors[10] and other attitudes, beliefs, and experiences that might affect medication nonadherence[11] might be potential causal pathways to disparities in chronic disease outcomes such as BP control.

Bosworth et al[12] proposed an organizing framework for the psychosocial and cultural domains they theorized would affect BP control, incorporating patient characteristics, including age, education, health literacy, and psychological factors such as beliefs and attitudes about health and illness, social/cultural environmental factors, including culturally linked perceptions of hypertension and therapies for it, and the medical environment, including the provider-patient relationship and interactions. The model does not specify causal associations, and, perhaps as a result, it has not been tested empirically, so the presence or strength of the hypothesized associations is not known. The model also did not include comorbid conditions. Thus, we sought to extend this work and the model itself by examining the contribution of the previously proposed and additional putative causal factors, in a more diverse sample in a different setting. We hypothesized that after adjusting for a more extensive set of potential confounders, race would no longer be significantly associated with BP control.

Received March 6, 2009; accepted December 3, 2009.

From the Section of General Internal Medicine (N.R.K., M.M.), Boston University School of Medicine, Boston, Mass; VA Boston Healthcare System (N.R.K.), Boston, Mass; Center for Health Quality, Outcomes, and Economic Research (N.R.K., M.B.O., M.E.G., D.B.), Bedford VAMC, Bedford, Mass; and the Health Policy and Management Department (M.E.G., D.B.), Boston University School of Public Health, Boston, Mass.

The views expressed in this article are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs.

Correspondence to Nancy Kressin, PhD, General Internal Medicine, Boston University School of Medicine, 801 Massachusetts Ave, No. 2090, Boston, MA 02118. E-mail nkressin@bu.edu

Circ Cardiovasc Qual Outcomes **is available at http://circoutcomes.ahajournals.org** **DOI: 10.1161/CIRCOUTCOMES.109.860841**

WHAT IS KNOWN

- Racial disparities in blood pressure (BP) control are well documented but poorly understood, and prior studies have only included a limited range of potential explanatory factors.

WHAT THE STUDY ADDS

- Blacks reported more experiences of discrimination, worse medication adherence, and more concerns about BP.
- Patient beliefs about hypertension can affect their BP control.
- Comorbid conditions, whose prevalence varies by race/ethnicity, can also affect BP control and thus are important to account for.
- We suggest an enhanced model of factors leading to disparities in BP control.

Methods

Sample

We identified all white and black patients ages 21 and older with 3 separate outpatient diagnoses of hypertension in 2004 in the primary care practices of a northeastern academically affiliated urban safety net tertiary care hospital. We use the term "black" to refer to patients of African or Caribbean descent. The study was approved by the institutional review board, and all subjects gave informed consent.

Study staff tracked clinic visits of these 10 125 patients over 19 months (October 2004 to June 2006), and, as they presented for care, approached 3526 of them to request study participation. Those willing were asked a series of questions and administered a 6-item cognitive screen to determine eligibility.[13] However, 654 patients (19% of 3526) overtly refused to participate and 920 patients (26% of 3526) responded that they were unable to participate that day but were potentially willing in the future, all before we were able to assess their eligibility. Subsequently, 1083 patients (55% of the remaining 1952) were excluded for reasons including cognitive difficulties, hearing impairment, not speaking English, or not being prescribed antihypertensive medications; we enrolled 869 patients. We then applied this 55% exclusion rate to the 1574 nonrespondents whose eligibility we had been unable to assess (654 refusers and 920 with no time) and estimated that 708 (45%) probably would have been eligible. Thus, we calculated our participation rate as participants/participants+likely eligible subjects (869/869+708=55%). We subsequently excluded 63 additional patients for whom study staff were unable to obtain BPs, for a final sample of 806 (Figure 1).

We evaluated the representativeness of the enrolled cohort, compared with those eligible to be enrolled, using the limited data available on the nonenrolled patients. Enrolled patients were more likely to be white (43% versus 32%, $P<0.0001$) and were younger (mean age, 59 versus 65 for nonenrolled patients, $P<0.0001$), but there was no difference in sex distribution from the parent population of eligible patients (not shown).

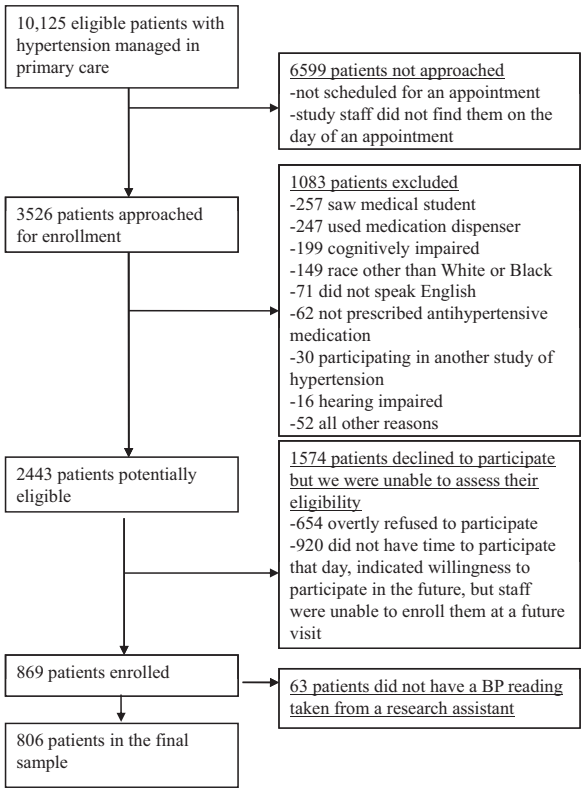


Figure 1. Flow chart of patient recruitment.

Measures

Patient Characteristics

Sociodemographic Characteristics and Medical History

Patient sociodemographic characteristics including race (assessed using US Census categories), education, and income were obtained through self-report. To assess patient literacy, we used the REALM-Short Form.[14] We used 5 separate dichotomous (yes/no) questions to assess insurance status, asking if patients currently have health insurance coverage through Medicare, Medicaid, Medigap, Free Care, or other insurance.

Patient clinical data were extracted from the electronic medical record, including age, sex, height, weight, and hypertension diagnosis. The electronic medical record was searched to obtain diagnoses of comorbid conditions including renal insufficiency, coronary artery disease, peripheral vascular disease, nicotine dependence, hyperlipidemia, diabetes mellitus, congestive heart failure, cerebrovascular disease, and obesity, because these conditions may influence the management of hypertension.[15] Obesity was considered a current diagnosis for any patient with an electronic medical record diagnosis or a calculated body mass index of at least 30.

Health Beliefs and Illness Perception

We examined a broad spectrum of patient beliefs and perceptions about high BP and related medications. We used the "Beliefs about Medicines Questionnaire"; 10 items assess patient concerns about present and potential future adverse effects from their medications and 8 items measure patients' beliefs regarding the necessity of their medications; we edited the items to relate responses specifically to BP medicines[16,17] (items scored on a 5-point scale: strongly agree to strongly disagree). Scores were summed within each scale to create an overall scale score (range, 8 to 40 for the "necessity" scale, Cronbach $\alpha=0.81$; 10 to 50 for the "concerns" scale, $\alpha=0.80$). Each score was divided by the number of items to obtain a mean summary score (range, 1 to 5); higher numbers indicated either greater concerns about medications or greater beliefs in their necessity.

We used 4 additional items from our prior work[18] to evaluate the degree of seriousness with which patients perceive hypertension, asking "How serious do you think high BP is, in general?"; "How serious do you think your high BP is, given your current use of medication?"; "If you did not take your BP medications, how likely do you think it would be that your BP would get worse over the next year?"; and "If you did not take your BP medications, how likely do you think it is that you would develop other health problems over the

next year?" (all scored on a scale of 1 to 5: extremely serious to not at all serious for the former 2 questions; very likely to very unlikely for the latter 2). We also included 10 individual items from the "cause" dimension of the "Illness Perception Questionnaire" to assess illness identity, cause, timeline, consequences, and cure control to examine patients' subjective beliefs about the etiology of their high BP (responses on a 5-point scale; range, strongly agree to strongly disagree).[19]

Perceived Discrimination in Health Care

We included 3 measures to assess patient perceptions of race-based discrimination in the health care setting. We used 5 questions that were a subset of the Commonwealth Fund 2001 Health Care Quality survey, focusing specifically on the patient perception of his/her provider and of encountering discrimination while receiving medical care, in general.[20] We created an additional question about the patient's perception of his/her provider's understanding of the patient's cultural background and how it affects his/her health; each item was examined individually. We also included 7 dichotomous items from a measure of patient perceptions of discrimination in accessing health care,[21] counting the number of experiences reported and creating a continuous variable (higher score indicates more experiences of discrimination[21]; α=0.90).

Medication Adherence and Hypertension Management

To assess medication adherence, we used the Hill-Bone Compliance to High Blood Pressure Therapy Scale, comprising items scored on a 4-point scale ("none of the time" to "all of the time").[22] We included the 9-item adherence subscale, which has been validated against BP control, summing the items to create a scale score (range, 9 to 36; higher scores indicate less adherence; α=0.74).

Outcome Assessment: BP Control

Research staff assessed patient BP using an automatic, portable machine (Omron HEM-907, Bannockburn, Ill), which was validated according to the international validation protocol and deemed an appropriate instrument for accurate BP measurement.[23] We excluded the 7% of patients without a BP reading from study staff. We dichotomized the BP readings to indicate whether each patients' BP was controlled or not, for example, when their systolic BP exceeded 140 mm Hg or their diastolic BP exceeded 90 mm Hg, according to the Joint National Committee on Hypertension 7 standards at the time of the study, which also specified that for patients with diabetes or renal insufficiency, BP should not exceed 130/80 mm Hg.[15]

Statistical Analyses

We first examined bivariate associations between race and each of the sociodemographic, clinical, attitudinal, experiential, and medication adherence variables, using χ^2 or t tests, as appropriate. Next, we performed bivariate analyses to determine whether this same set of variables was associated with BP control (yes/no). Finally, BP control was modeled as a function of the covariates using random-effects logistic regression. The random effects, which were assumed to be mean-zero gaussian additive errors on the logit scale, accounted for 2 levels of clustering: patients within providers and providers within clinics.

We fit 6 sets of models of increasing numbers of covariates, including only patients with complete data. In the first, we examined the effects of race alone on BP control. Next, we examined the effects of race on BP control, adjusting for age and other sociodemographic characteristics. The third model added medication adherence, the fourth added health beliefs, the fifth added experiences of discrimination, and the sixth added comorbid conditions. Within each model, we chose the subset of additional variables through an ordinary logistic regression stepwise selection procedure, forcing in provider indicators as fixed effects, keeping significant variables from the prior model, and including race in all models. This procedure prevents collinearity and overparameterization. We used a probability value of 0.05 for variables to enter or be removed from the model. *c*-Statistics and Hosmer-Lemeshow analyses were performed on the models resulting from the stepwise procedure. Each model was then rerun with the random effects terms. All analyses were conducted using SAS 9.1.3 (SAS Institute, Cary, NC) with the exception of the random-effects logistic regression, which was conducted using the statistics package *R* (version 2.8.0[24]).

Results

Sociodemographic Characteristics of the Sample

Most of the sample were black (57%) or female (65%), with an overall mean age of 59 years. White patients were older (61 versus 58 years), and more likely to have at least a high school education (90% versus 71%). Black patients were more likely to have insurance coverage through Medicaid or Free Care (42% versus 22%, 41% versus 25%, respectively), whereas white patients were more likely to have "other" (private) insurance (59% versus 35%). There were race differences in combined household family income, with a greater percentage of blacks earning low incomes of <$20 000 (58% versus 36% whites) and being less likely to have a literacy score of 9th grade or higher (48% versus 84%), (all P<0.001; Table 1).

BP and Medical History by Race

A greater percentage of white patients had controlled BP (65% versus 53%), with lower systolic and diastolic BP than blacks (white systolic BP, 130 mm Hg versus 132 mm Hg; white diastolic BP 76 mm Hg versus 81 mm Hg). White patients were more likely to have hyperlipidemia (60% versus 47%), peripheral vascular disease (7% versus 4%), benign prostatic hypertrophy (6% versus 2%), coronary artery disease (19% versus 8%), and cerebrovascular disease (7% versus 4%). Black patients were more likely to have diabetes (40% versus 24%), renal insufficiency (7% versus 4%), congestive heart failure (5% versus 1%), and to be obese (65% versus 52%, all P<0.05).

Health Beliefs and Illness Perceptions

Blacks had significantly more concerns about their BP medications than whites (2.5 versus 2.1, P<0.0001). White patients were significantly more likely to respond that their BP was less serious, given their current use of medications (3.3 versus 2.8, P<0.0001).

When asked about the causality of high BP, blacks agreed more with the notion that it is caused by a germ or virus or that diet, pollution, or heredity played a major role in causing BP (Table 1). Blacks were more likely to indicate that it was just by chance that they became ill with hypertension, that other people played a large role in causing their BP, or that high BP was caused by poor medical care in the past.

Perceived Discrimination

When asked if there was ever a time they would have gotten better medical care if they belonged to a different race or ethnic group or if they ever felt that a doctor or medical staff judged them unfairly or treated them with disrespect because of how well they spoke English, blacks were more likely to respond "yes" than whites (19% versus 1% and 4% versus 0%, respectively). Although all patients generally agreed that their provider understood their background and values, black patients agreed less strongly (1.5 versus 1.4), and though all patients disagreed that their provider looks down on them and

Table 1. Sociodemographic, Clinical, Attitudinal, Belief, and Experience Variables by Race and BP Control

Sociodemographic Characteristics and Medical History	Overall, %	White, %	Black, %	*P* Value for Race Differences	Controlled BP, %	Uncontrolled BP, %	*P* Value for Difference by BP Control
% White	**43**			**n/a**	**48**§	**36**§	**0.0009**
Mean age, y	**59**	**61**	**58**	**0.0007**	59	60	0.0853
% Male	**35**	**46**	**27**	**<0.0001**	34	38	0.2919
Education, % ≥12 y	**79**	**90**	**71**	**<0.0001**	81	76	0.0521
% Income <$20 000	**48**	**36**	**58**	**<0.0001**	**43**	**56**	**0.0003**
Insurance status							
Medicare	39	39	40	0.6889	37	43	0.0984
Medigap	3	3	2	0.4991	2	4	0.2143
Medicaid	**33**	**22**	**42**	**<0.0001**	**29**	**40**	**0.0012**
Other	**45**	**59**	**35**	**<0.0001**	**51**	**38**	**0.0003**
Free care	**34**	**25**	**41**	**<0.0001**	32	37	0.1072
Literacy categories							
≤3rd grade	**4**	**2**	**5**		3	4	
4th to 6th grade	**9**	**2**	**14**		8	11	
7th to 8th grade	**24**	**12**	**33**		22	27	
≥9th grade	**63**	**84**	**48**	**<0.0001**	67	58	0.0884
% With controlled BP	**58**	**65**	**53**	**0.0009**	n/a	n/a	n/a
Mean systolic BP, mm Hg	**131**	**130**	**132**	**0.0277**	n/a	n/a	n/a
Mean diastolic BP, mm Hg	**79**	**76**	**81**	**<0.0001**	n/a	n/a	n/a
Nicotine dependence	7	5	9	0.0839	7	8	0.7303
Hyperlipidemia	**52**	**60**	**47**	**0.0002**	52	53	0.9302
Diabetes	**33**	**24**	**40**	**<0.0001**	**25**	**45**	**<0.0001**
Peripheral vascular disease	**5**	**7**	**4**	**0.0383**	5	5	0.9918
Renal insufficiency	**6**	**4**	**7**	**0.0470**	**3**	**10**	**<0.0001**
Benign prostatic hypertrophy	**4**	**6**	**2**	**0.0003**	4	2	0.1107
Coronary artery disease	**13**	**19**	**8**	**<0.0001**	14	11	0.2151
Obesity	**59**	**52**	**65**	**0.0004**	58	62	0.2373
Congestive heart failure	**3**	**1**	**5**	**0.0091**	2	5	0.0635
Cerebrovascular disease	**5**	**7**	**4**	**0.0382**	4	6	0.2162
Health Beliefs and Illness Perceptions							
BMQ: Mean necessity of medications, mean score	3.7	3.7	3.7	0.1542	3.7	3.7	0.9739
BMQ: Mean concerns about medications, mean score	**2.3**	**2.1**	**2.5**	**<0.0001**	**2.3**	**2.4**	**0.0407**
How serious do you think high BP is, in general*	1.5	1.5	1.5	0.4607	1.5	1.6	0.1147
How serious do you think your high BP is, given your current use of medication?*	**3.0**	**3.3**	**2.8**	**<0.0001**	**3.1**	**2.9**	**0.0010**
If you did not take BP meds, likelihood that BP would get worse w/in a year†	1.5	1.4	1.5	0.1284	1.5	1.4	0.1348
If you did not take BP meds, likelihood that you would develop other health problems w/in a year†	1.7	1.7	1.7	0.2307	1.7	1.7	0.4281
Illness Perceptions Questionnaire Items‡							
A germ or virus caused my high BP	**3.9**	**4.2**	**3.7**	**<0.0001**	3.9	3.8	0.1208
Diet played a major role in causing my high BP	**2.3**	**2.4**	**2.1**	**0.0017**	2.2	2.3	0.6715
Pollution caused my high BP	**3.7**	**3.8**	**3.5**	**0.0002**	3.7	3.7	0.7546

(*Continued*)

Table 1. Continued

	Overall	White, %	Black, %	*P* Value for Race Differences	Controlled BP, %	Uncontrolled BP, %	*P* Value for Difference by BP Control
My high BP is hereditary—it runs in my family	**2.0**	**2.1**	**1.9**	**0.0134**	2.1	2.0	0.1877
It was just by chance that I became ill with high BP	**3.5**	**3.7**	**3.3**	**<0.0001**	3.5	3.4	0.1876
Stress was a major factor in causing my high BP	2.4	2.4	2.4	0.4035	2.4	2.4	0.4737
My high BP is largely due to my own behavior	2.7	2.7	2.8	0.3299	2.8	2.7	0.1508
Other people played a large role in causing my high BP	**3.4**	**3.6**	**3.3**	**0.0010**	3.4	3.3	0.2580
My high BP was caused by poor medical care in the past	**3.9**	**4.1**	**3.7**	**<0.0001**	3.9	3.8	0.2799
My state of mind played a major part in causing my high BP	3.2	3.3	3.2	0.7686	3.3	3.2	0.4884
Perceived Discrimination							
Commonwealth Fund Items							
Was there ever a time you would have gotten better medical care if you belonged to a different race or ethnic group? (% yes)	**11**	**1**	**19**	**<0.0001**	12	11	0.6174
In the last 2 years, have you ever felt that the doctor or medical staff judged you unfairly or treated you with disrespect because of how well you speak English? (% yes)	**2**	**0**	**4**	**0.0008**	3	2	0.3549
My provider treats me with a great deal of respect and dignity‡	1.3	1.3	1.3	0.2509	1.3	1.3	0.5617
I feel that my provider understands my background and values‡	**1.5**	**1.4**	**1.5**	**0.0425**	1.5	1.5	0.1818
I often feel as if my provider looks down on me and the way I live my life†	**4.4**	**4.5**	**4.3**	**<0.0001**	4.4	4.4	0.7828
New item							
I feel my provider understands my cultural background and how it affects my health‡	1.8	1.8	1.8	0.5104	1.8	1.8	0.1044
Perceived Discrimination Scale (Bird and Bogart)							
Discrimination scale	**0.7**	**0.2**	**1.1**	**<0.0001**	0.7	0.7	0.9815
Medication adherence	**10.5**	**9.9**	**11.0**	**<0.0001**	**10.4**	**10.7**	**0.0347**

Bolded text indicates significant differences; because of rounding, items sum to more than 100%.
*Higher score indicates greater belief that high BP is not serious.
†Higher score indicates greater belief that statement is unlikely to be true.
‡Higher score indicates more disagreement with statement.
Mean necessity: scale, 1 to 5. Higher score indicates medications are a necessity.
Mean concerns: scale, 1 to 5. Higher score indicates greater concern about medications.
Perceived discrimination scale: Higher score indicates more experiences of discrimination.
Hill Bone Adherence: Higher score indicates less adherence.
§Percent white.

the way they live their life, white patients disagreed more strongly (4.5 versus 4.3). Blacks reported more experiences of discrimination when receiving health care than did whites (1.1 versus 0.2, all $P<0.05$).

Medication Adherence

White patients reported better medication adherence than did black patients (9.9 versus 11.0, $P<0.0001$).

Covariates by BP Control

A greater proportion of patients with controlled BP were white compared with patients with uncontrolled BP (48% versus 36%). Patients with uncontrolled BP were more likely to have a household income of <$20 000 (56% versus 43%), were more likely to have Medicaid (40% versus 29%), and were less likely to have "other" insurance (38% versus 51%). Patients with uncontrolled BP were more likely to have

Table 2. Multivariate Results Modeling Controlled BP

	Model 1		Models 2 and 3		Model 4		Model 5		Model 6	
	Coefficient	Lower CI, Upper CI†	Coefficient	Lower CI, Upper CI†	Coefficient	Lower CI, Upper CI†	Coefficient	Lower CI, Upper CI†	Coefficient	Lower CI, Upper CI†
Race (white)	0.42*	0.13, 0.71	0.37*	0.06, 0.68	0.33	−0.003, 0.66	0.48*	0.13, 0.83	0.32	−0.03, 0.67
Income			−0.44	−0.75, −0.13	−0.37	−0.68, −0.06	−0.37	−0.70, −0.04	−0.32	−0.65, 0.01
My high BP is largely due to my own behavior‡					0.15	0.01, 0.29	0.17	0.03, 0.31	0.16	0.02, 0.30
How serious do you think high BP is, in general?					−0.23	−0.43, −0.03	−0.24	−0.44, −0.04	−0.23	−0.43, −0.03
How serious do you think your high BP is, given your current use of medication?					0.18	0.04, 0.32	0.18	0.04, 0.32	0.17	0.01, 0.33
Would have gotten better medical care if you belonged to different race or ethnic group							0.56	0.05, 1.07	0.51	−0.02, 1.04
Diabetes									−0.62	−0.97, −0.27
Renal insufficiency									−1.17	−1.93, −0.41
Benign prostatic hypertrophy									0.93	−0.13, 1.99

*$P \leq 0.05$ (value only reported for race, the only variable forced into all of the stepwise selection procedures).
†95% confidence intervals.
‡Higher score indicates more disagreement with statement.

diabetes (45% versus 25%) and renal insufficiency (10% versus 3%) and concerns about their BP medications (2.4 versus 2.3). Patients with controlled BP were more likely to disagree that their own BP was serious, given their current use of medication (3.1 versus 2.9), and reported better medication adherence (10.4 versus 10.7, all $P<0.05$).

Multivariate Results

The first model, only adjusted for race, accounting for physician and clinic, indicated that white patients had higher odds of having controlled BP than blacks (model 1, b=0.42, $P=0.0068$; Table 2; *c*-statistic (*c*)=0.661; Hosmer and Lemeshow goodness-of-fit test [HL] $P=0.4226$). The effect of race on BP control persisted in the second model, after adjustment for income (other sociodemographic variables were excluded through the ordinary logistic regression stepwise procedure). The third model added a measure of adherence, which was not significantly related to BP control, so these 2 latter models had the same results, with race continuing to be a significant predictor of BP control (models 2 and 3, b=0.37, $P=0.0238$; model 2, $c=0.691$, HL $P=0.7127$; model 3, $c=0.670$, HL $P=0.9260$). In the fourth model, an item assessing patients' beliefs that high BP is largely due to one's own behavior was added, as well as 2 items assessing the degree of seriousness with which patients perceive hypertension (all other attitudinal and experiential factors were excluded through the stepwise procedure). In this model, race was no longer significantly associated with odds of BP control (model 4, b=0.33, $P=0.0531$, $c=0.713$, HL $P=0.3932$).

In the fifth model, 1 item assessing experiences with perceived discrimination in health care was added, specifically, perceptions about whether they would have ever gotten better medical care if they belonged to a different race or ethnic group (other items assessing discrimination were excluded through the stepwise procedure). Race was significant in this model (model 5, b=0.48, $P=0.0074$, $c=0.708$, HL $P=0.8660$).

In the final model, diabetes, renal insufficiency, and benign prostatic hypertrophy were added (other comorbid conditions were excluded during the stepwise procedure in SAS). Here, race was no longer a significant predictor of BP control (model 6, b=0.32, $P=0.0876$, $c=0.740$, HL $P=0.4274$).

Discussion

Understanding and ameliorating racial disparities in BP control is a major public health and clinical concern. We hypothesized that after adjusting for an extensive set of potential confounders, race would no longer be significantly associated with BP control, and the results generally supported this notion. Although the effects of race persisted after accounting for sociodemographic factors, the inclusion of BP-related attitudes and beliefs rendered race insignificant. However, the introduction of the discrimination variables made race significant again, in a counterintuitive fashion, although in the final model, with the inclusion of comorbid conditions, race was no longer significant. The finding that

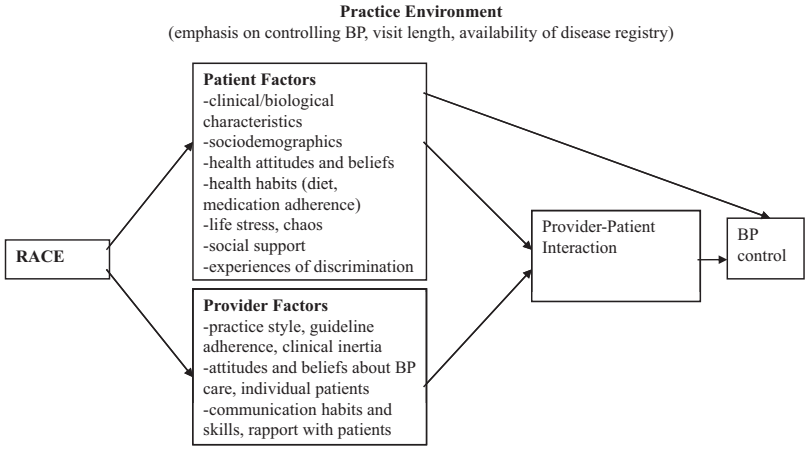


Figure 2. Expanded model of factors leading to disparities in BP control.

patients who agreed with the statement that there was "ever a time when they would have gotten better medical care if they had belonged to a different race/ethnic group" had better BP control is puzzling. We carefully explored this dynamic, ruling out the possibility that race-discrimination interactions were driving it, and finding that even if we removed this variable, another discrimination variable became significant in the same counterintuitive direction (not shown). The variable was limited by its reference point (eg, was there ever a time . . .), so it is possible that patients who felt they were getting bad care in the past had changed clinicians and are now getting good care, leading to better current BP care and control. It is also possible that another, unmeasured, confounder caused these results.

The study findings are different than the Bosworth study,[11] which indicated that after controlling for a similar range of factors (not including discrimination or comorbid conditions), race remained a significant predictor of BP control, among VA patients in one Southern city. They speculated that the results they obtained in that setting, where access to care is assured, might be different than those found elsewhere, and our results in a northeastern urban safety net hospital setting support this notion. Several other studies also controlled for subsets of the factors we included here, but race remained significant.[6,25,26] Thus, although our findings support Bosworth's proposed framework as a descriptive model, and the notion that numerous psychosocial, behavioral, and experiential factors mediate the relationship between race and BP control,[11] they indicate that comorbid conditions, whose prevalence varies by race/ethnicity, are also important to account for in models of disparities in BP control.

Several findings have clinical implications. Blacks reported more experiences of discrimination, and such experiences may erode overall trust in physicians, their diagnoses, and the therapies they prescribe. Experiences of discrimination in the community setting are generally associated with higher BP[27,29,30] with less use of chronic disease care[28] and may negatively affect patients' acceptance of their diagnosis and beliefs in the necessity of or concerns about the associated therapy, which are foundational to patient adherence to prescribed medications.[29] Unequal treatment, documented by us and others,[30,31] may also contribute to disparities in BP outcomes.

Blacks indicated worse medication adherence and more concerns about BP. Each of these is potentially ameliorable through educational or counseling interventions, and our results suggest that addressing these will help address disparities in BP control.

This study was limited by its focus on patients in a single setting and its inclusion of only black (albeit both African-American and Caribbean-born) and white patients. Although we required that participants have 3 separate outpatient diagnoses of hypertension, our inability to contact or enroll many eligible patients may have biased our sample toward more frequent users. The observational nature of these data limits our ability to form causal inferences because we were not able to randomize by attitudinal characteristics or ascertain that certain attitudes or experiences preceded BP outcomes in time. Further, our measure of adherence, although internally consistent and previously validated against BP control, was obtained by self-report. However, the large sample, which included women and detailed assessments of the richest array of putative factors examined to date, offsets the limitations.

Among the potential causes of disparities in BP control, the etiologic factors could arise from the patient (health beliefs and experiences, medication adherence or self-care behaviors, clinical or biological factors), the provider (practice style, communication skills, attitudes), the doctor-patient interaction, or the environment. Although the Bosworth model provided an excellent summary of the psychosocial and cultural factors that might be associated with BP control, we propose that a model for racial disparities in BP control should include a wider array of factors and include hypothesized associations (Figure 2). The present results help to demonstrate the effects of a variety of factors on race differences in BP, but we lacked data on other self-care behaviors important to hypertension management (eg, diet, exercise) to fully address this question. Nor are we able to rule out biological differences, such as race-linked nitric oxide deficiencies associated with cardiovascular disease[32] or differences in the process of care. Our future work will also examine the effects of racial differences in providers' therapeutic intensification, which varies by race,[33] on BP outcomes. Similar to our prior suggestions for future directions in research on racial differences in invasive cardiac procedure

use,[31] here we propose additional careful attention by clinicians, researchers, and, ultimately, policymakers, to a comprehensive array of factors associated with race to identify and intervene on the causes of racial disparities in BP outcomes.

Acknowledgments

We thank Peter Davidson, MD, and the clinic and research staff for their assistance.

Sources of Funding

This study was supported by NIH/National Heart, Lung, and Blood Institute grant R01 HL072814 (Dr Kressin, PI); Dr Kressin is also supported by a Research Career Scientist award (RCS 02-066-1) from the Health Services Research and Development Service, Department of Veterans Affairs.

Disclosures

None.

References

1. American Heart Association. Heart *Disease and Stroke Statistics: 2009 Update At-A-Glance*. Available at: http://www.americanheart.org/downloadable/heart/1240250946756LS-1982%20Heart%20and%20Stroke%20Update.042009.pdf. Accessed November 20, 2009.
2. Wang X, Poole J, Treiber F, Harshfield, GA, Hanevold C, Snieder H. Ethnic and gender differences in ambulatory blood pressure trajectories: results from a 15-year longitudinal study in youth and young adults. *Circulation*. 2006;114:2780–2787.
3. Wong M, Shapiro M, Boscardin W, Ettner S. Contributions of major diseases to disparities in mortality. *N Engl J Med*. 2002;347:1585–1592.
4. Cushman W, Ford C, Cutler J, Margolis K, Davis B, Grimm R, Black H, Hamilton B, Holland J, Nwachuku C, Papademetriou V, Probstfield J, Wright J, Alderman M, Weiss R, Piller L, Bettencourt J, Walsh W. Success and predictors of blood pressure control in diverse north am settings: the Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT). *J Clin Hypertens*. 2002;4:393–404.
5. Hyman DJ, Pavlik VN. Self-reported hypertension treatment practices among primary care physicians. *Arch Intern Med*. 2000;160:2281–2286.
6. Hajjar I, Kotchen T. Trends in prevalence, awareness, treatment, and control of hypertension in the United States, 1988–2000. *JAMA*. 2003;290:199–206.
7. Hicks LS, Fairchild DG, Horng MS, Orav EJ, Bates DW, Ayanian JZ. Determinants of JNC VI guideline adherence, intensity of drug therapy, and blood pressure control by race and ethnicity. *Hypertension*. 2004;44:429–434.
8. Fletcher SW, Deliakis J, Schoch WA, Shapiro SH. Predicting blood pressure control in hypertensive patients: an approach to quality-of-care assessment. *Med Care*. 1979;27:285–292.
9. Wagner EH, James SA, Beresford SA, Strogatz DS, Grimson RC, Kleinbaum DG, Williams CA, Cutchin LM, Ibrahim MA. The Edgecombe County High Blood Pressure Control Program, I: correlates of uncontrolled hypertension at baseline. *Am J Public Health*. 1984;74:237–242.
10. Heisler M, Faul JD, Hayward RA, Langa KM, Blaum C, Weir D. Mechanisms for racial and ethnic disparities in glycemic control in middle-aged and older Americans in the health and retirement study. *Arch Intern Med*. 2007;167:1853–1860.
11. Bosworth HB, Dudley T, Olsen MK, Voils CI, Powers B, Goldstein MK, Oddone EZ. Racial differences in blood pressure control: potential explanatory factors. *Am J Med.* 2006;119:70.e9–e15.
12. Bosworth H, Oddone E. A model of psychosocial and cultural antecedents of blood pressure control. *J Natl Med Assoc*. 2002;94:236–248.
13. Callahan CM, Unverzagt FW, Hui SL, Perkins AJ, Hendrie HC. Six-item screener to identify cognitive impairment among potential subjects for clinical research. *Med Care*. 2002;40:771–781.
14. Arozullah AM, Yarnold PR, Bennett CL, Soltysik RC, Wolf MS, Ferreira RM, Lee S-Y, Costello S, Shakir A, Denwood C, Bryant FB, David T. Development and validation of a short-form, rapid estimate of adult literacy in medicine. *Med Care*. 2007;45:1026–1033.
15. Chobanian A, Bakris G, Black H, Cushman W, Green L, Izzo J Jr, Jones D, Materson B, Oparil S, Wright J Jr, Rocella E, and the 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.
16. Horne R, Weinman J, Hankins M. The Beliefs about Medicines Questionnaire: the development and evaluation of a new method for assessing the cognitive representation of medication. *Psychol Health*. 1999;14:1–24.
17. Horne R, Buick D, Fischer M, Leake H, Cooper V, Weinman J. Doubts about the necessity and concerns about adverse effects: identifying the types of beliefs that are associated with non-adherence to HAART. *Int J STD AIDS*. 2004;15:38–44.
18. Kressin N, Wang F, Long J, Bokhour B, Orner M, Rothendler J, Clark C, Reddy S, Kozak W, Kroupa L, Berlowitz D. Hypertensive patients' health beliefs, process of care, and medication adherence: is race important? *J Gen Intern Med*. 2007;22:768–774.
19. Weinman J, Petrie KJ, Moss-Mossis R, Horne R. The Illness Perception Questionnaire: a new method for assessing the cognitive representation of illness. *Psychol Health*. 1996;11:431–445.
20. Johnson RL, Saha S, Arbelaez JJ, Beach MC, Cooper LA. Racial and ethnic differences in patient perceptions of bias and cultural competence in health care. *J Gen Intern Med*. 2004;19:101–110.
21. Bird ST, Bogart LM. Perceived race-based and socioeconomic status (SES)-based discrimination in interactions with health care providers. *Ethnic Dis*. 2001;11:554–563.
22. Kim MT, Hill MN, Bone LR, Levine DM. Development and testing of the Hill-Bone Compliance to High Blood Pressure Therapy scale. *Prog Cardiovasc Nursing*. 2000;15:90–96.
23. El Assaad MA, Topouchian JA, Darne BM, Asmar RG. Validation of the Omron HEM-907 device for blood pressure measurement. *Blood Press Monit*. 2002;7:237–241.
24. *R: A Language Environment for Statistical Computing* [computer program]. Vienna, Austria: R Foundation for Statistical Computing; 2008.
25. Kramer H, Han C, Post W, Goff D, Diez-Roux A, Cooper R, Jinagouda S, Shea S. Racial/ethnic differences in hypertension and hypertension treatment and control in the multi-ethnic study of atherosclerosis (MESA). *Am J Hypertens*. 2004;17:963–970.
26. Svetkey LP, George LK, Tyroler HA, Timmons PZ, Burchett BM, Blazer DG. Effects of gender and ethnic group on blood pressure control in the elderly. *Am J Hypertens*. 1996;9:529–535.
27. Krieger N, Sydney S. Racial discrimination and blood pressure: the CARDIA study of young black and white adults. *Am J Public Health*. 1996;86:1370–1378.
28. Blanchard J, Lurie N. R-E-S-P-E-C-T: patient reports of disrespect in the health care setting and its impact on care. *J Fam Pract*. 2004;53:721–730.
29. Rosenstock I. Why people use health services. *Milbank Memorial Fund Q*. 1966;44.
30. Kressin N. 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.
31. Institute of Medicine. *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care.* Washington, DC: National Academy Press; 2002.
32. Kalinowski L, Dobrucki IT, Malinski T. Race-specific differences in endothelial function: predisposition of African Americans to vascular diseases. *Circulation*. 2004;109:2511–2517.
33. Manze M, Rose AJ, Orner MB, Berlowitz D, Kressin NR. Understanding pathways to disparities in hypertension control: race and treatment intensification. Miami, FL: Society of General Internal Medicine; 2009.

ELSEVIER

Women's Health Issues 20 (2010) 105–113

WOMEN'S HEALTH ISSUES

www.whijournal.com

NURSING HOME RESIDENCE CONFOUNDS GENDER DIFFERENCES IN MEDICARE UTILIZATION
An Example of Simpson's Paradox

Andrea C. Kronman, MD, MSc[a,c,*], Karen M. Freund, MD, MPH[a,c], Amresh Hanchate, PhD[a], Ezekiel J. Emanuel, MD, PhD[b], and Arlene S. Ash, PhD[a]

[a]*Women's Health and Health Care Research Units, Section of General Internal Medicine, Evans Department of Medicine, Boston University Medical Center, Boston, Massachusetts*

[b]*Department of Clinical Bioethics, Warren Magnuson Clinical Center, National Institutes of Health, Bethesda, Maryland*

[c]*Women's Health Interdisciplinary Research Center, Boston University School of Medicine, Boston, Massachusetts*

Received 26 December 2007; revised 1 August 2009; accepted 21 November 2009

Background. **Gender differences in health care utilization in older Americans may be confounded by nursing home residence. Medicare data contain several files that can be used to create a measure of nursing home residence, but prior work has not addressed which best account for potential confounding. Simpson's paradox occurs when aggregated data support a different conclusion from what the disaggregated data show. We describe such a paradox that appeared when we sharpened our definition of "nursing home residence" while examining gender differences in Medicare utilization at the end of life.**

Methods. **To understand gender-specific health care utilization at the end of life, we conducted a retrospective analysis of a national random sample of Medicare beneficiaries aged 66 or older who died in 2001 with Parts A and B data for 18 months before death. We sought to associate each of total hospital days and costs during the final 6 months of life with numbers of primary care physician visits in the 12 preceding months. In addition to demographics, comorbidities, and geography, "nursing home residence" was a potential confounder, which we imputed in two ways: 1) from skilled nursing facility bills in the Part A Medicare Provider Analysis and Review (MedPAR) file; and 2) from Berenson-Eggers-Type-of-Service codes indicating widely spaced doctor visits in nursing homes obtained from Medicare's carrier file.**

Conclusion. **Gender differences in Medicare utilization are strongly confounded by nursing home resident status, which can be imputed well from Medicare's carrier file, but not MedPAR.**

Background

During this time of health care reform, understanding gender differences in health care utilization in the Medicare population is critical, because women live longer and make up a larger proportion of the Medicare population (Medicare Beneficiary Demographics, 2006). In addition, women live with greater disability and more chronic disease than men. For example, chronic conditions such as diabetes, arthritis, dementia, depression, and obesity are more common in women (Lubitz & Riley, 1993). As baby boomers age, Medicare faces escalating expenditures and dwindling numbers of workers to fund it. Because 30% of Medicare expenditures are spent on the 6% of beneficiaries who die each year (*Approaching Death*, 1997; Edwards & DeHaven, 2003; Hogan, Lunney, Gabel, & Lynn, 2001), there is strong interest in improving the efficiency of health care at the end of life. Primary care medicine may improve this efficiency, because it is intended to address comprehensiveness, coordination, continuity, and "sustained clinician–patient partnerships" with the family and community (Montgomery

Funded by the Office of Research on Women's Health (ORWH) K12-43444, with support from the Centers for Medicare and Medicaid Services (CMS) and the National Cancer Institute (NCI).

* Correspondence to: Andrea C. Kronman, MD, MSc, Women's Health Unit, 801 Massachusetts Avenue, Suite 470, Boston, MA 02118; Phone: 617-638-8036; Fax: 617-638-8096.

E-mail: Andrea.Kronman@bmc.org.

1049-3867/10 $-See front matter.
doi:10.1016/j.whi.2009.11.016

et al., 2004). In addition, primary care providers may focus—more than specialists—on preventing medical complications, discussing patient preferences, and coordinating home palliative care at life's end. Our prior work found more primary care visits in the preceding year were associated with reduced hospitalization, in-hospital death, and costs in the last 6 months of life (Kronman, Ash, Freund, Hanchate, & Emanuel, 2008).

We hypothesized that there were gender differences in the relationship between primary care and health care utilization at the end of life. Women tend to use more primary care and outpatient services, whereas men tend to use more in-patient hospital services (Bird, Shugarman, & Lynn, 2002; Song, Chang, Manheim, & Dunlop, 2006). Because Medicare currently rewards in-hospital services more than primary care, the program's financial incentives may affect men and women differently. To understand gender differences in the health care utilization of an elderly population, it is important to adjust for those who are nursing home residents, because more women are residents than men, and residents have different utilization of both primary care and hospital services. Because Medicare data do not contain a reliable indicator of nursing home resident status, we report here on two different methods of imputing nursing home resident status and demonstrate the effect on health care utilization outcomes.

Methods

Data source

To examine gender differences in end-of-life care, we used a randomly sampled population of 116,318 Medicare beneficiaries aged 66 or older who died in the last 6 months of 2001. Non-Whites were oversampled because the study population had been constructed to focus on end-of-life health care disparities (Hanchate, Kronman, Young-Xu, Ash, & Emanuel, 2009). To ensure completeness and comparability of health care utilization records before death, we excluded people who were not continuously enrolled in Medicare Parts A and B traditional fee-for-service program for the final 18 months of life. We also excluded decedents who could not be matched to the National Death Index, and those enrolled in the end-stage renal disease program, leaving a final analytical sample of 78,353.

Measures

We obtained age, gender, race, and zip code of residence from the Medicare denominator file, using the Medicare racial/ethnic categories of White, Black, Hispanic, and other (for those of Asian, North American Native, and other or unknown races and ethnicities). We used an indicator that Part B Medicare insurance had been purchased by a state in 2000 (state Medicaid "buy-in") as a proxy for low income. Place of death was derived from death certificates through the National Death Index. A summary comorbidity measure was determined using DxCG's prospective relative risk score (DCG version 6.1 for Windows [DxCG, 2004]), derived from the presence in the pre-period of ICD-9-CM diagnosis codes from inpatient and outpatient encounters in Medicare's utilization files. These encounters include all physician visits, hospital care, and nursing home care, but not codes used for diagnostic tests. The score is calibrated to associate 1.0 with average expected expenditures in the following year among all Medicare beneficiaries observed during routine 12-month periods (Ash et al., 2000).

Because Medicare Part A only pays for nursing home visits in the first 90 days after a hospitalization, we were unable to directly distinguish long-term nursing home residents from those in a nursing home for short-term rehabilitation after a hospitalization from our administrative billing data. We thus imputed two proxy measures. For the first measure, we defined those with "nursing home use" to be beneficiaries with at least one bill for residential costs in a skilled nursing facility (SNF) in the Part A Medicare Provider Analysis and Review (MedPAR) SNF file. This method is limited in not being able to distinguish nursing home residents from patients who are receiving short-term rehabilitation in a nursing facility. However, because MedPAR is more commonly utilized in health services research, this has the potential to be a readily utilized methodology. The second method utilizes the Medicare Part B carrier file. Unlike Medicare Part A, Medicare Part B does not pay for residential costs in a nursing home; however, it does pay for diagnostic tests and medical provider visits that take place in a nursing home. We used Berenson-Eggers-Type-Of-Service (BETOS) code 4B in Medicare's "Carrier" file (available from: http://www.cms.hhs.gov/HCPCSReleaseCodeSets/20_BETOS.asp) to define our second measure. This method of identifying nursing home use has been previously validated (Campbell et al., 2004). With it, a beneficiary was flagged for nursing home use if there were at least two health care encounters (e.g., blood draws and/or physician visits), separated by at least 90 days, that occurred in a nursing home. Because the maximum period of Medicare reimbursement for SNF care after a hospitalization is 90 days, people who received services spanning a longer period are likely to have resided in a nursing home receiving long-term care; although this method will also (spuriously) find occasional non-nursing home residents with rehabilitative SNF stays after two distinct hospitalizations.

Primary care visits

We used BETOS codes in the Medicare Carrier file to identify a visit in a nursing facility (code 4B) or office

(codes 1A and 1B). We then linked these codes to the Medicare Health Care Financing Administration specialty codes to define a visit to an internist (11), geriatrician (38), or family practitioner (08), as a "primary care visit." We used the number of these visits in the 12-month "pre-period" before the final 6 months of life to form five primary care groups: no visits, one or two, three to five, six to eight, and nine or more visits. We categorized visits that did not have these primary care Health Care Financing Administration specialty codes as "visits to specialists." We calculated the ratio of number of specialty visits to number of total outpatient encounters for each gender and by nursing home use.

Outcomes

We studied two outcomes during the last 6 months of life: 1) number of inpatient days (obtained from the MedPAR files) and 2) total costs paid by Medicare (from the MedPAR, Carrier, Durable Medical Equipment, Hospice, and Outpatient files).

Statistical analysis

We used bivariate analyses (chi-square test for categorical variables and analysis of variance for continuous variables) to identify differences in characteristics, end-of-life utilization, and costs across the primary care groups and genders. To compare the characteristics of our two definitions of nursing home use, we conducted bivariate analysis on the population within four groups, those with 1) both SNF bills from MEDPAR and encounters from carrier file, 2) only SNF bills, 3) only carrier file encounters, and 4) neither. STATA software version 9.1 (StataCorp, College Station, TX) was used for all analyses. We adjusted for factors from the pre-period that affect health care utilization and outcomes: age (Levinsky et al., 2001), race (Barnato, Lucas, Staiger, Wennberg, & Chandra, 2005; Degenholtz, Thomas, & Miller, 2003; Skinner, Weinstein, Sporer, & Wennberg, 2003), Medicaid buy-in, comorbidity (Ash et al., 2003; Hogan et al., 2001; Institute of Medicine, 1997), and place of residence (Baicker, Chandra, Skinner, & Wennberg, 2004; Fisher et al., 2000; Pritchard et al., 1998). For our first analysis, we adjusted for all confounders excluding either nursing home measure, which resulted in aggregated data around the confounder of nursing home use. To disaggregate the data, we then adjusted with the MedPAR-based measure of nursing home use, and a third time substituting the carrier file–based measure instead of the MedPAR measure. Our final analysis stratified the population by both gender and nursing home use based on the carrier file measure.

To account for geographic differences, we used multivariable cluster analysis, specifically, fixed effects regression with dichotomous indicators for each geographic unit. Contrasting each outcome only for beneficiaries residing in the same geographic area accounts for both measured and unmeasured factors (including health care supply and labor), which vary by geographic location (Johnston & DiNardo, 1997). We mapped each beneficiary's zip code of residence into its "hospital service area," defined by the Dartmouth as a geographic unit associated with health care utilization patterns (Fisher & Wennberg, 2008).

We calculated risk-adjusted, expected outcomes for each gender by primary care visit group by using its coefficient together with mean values for each of the other covariates in the model, using sample weights to reflect the entire Medicare population, and tested for statistical differences by gender. The key independent variables of our first three analyses were fully interacted 2×5 matrices of gender by number of pre-period primary care visits. Our final analyses used a fully interacted 4×5 matrix of gender and nursing home use by number of pre-period primary care visits.

Results

The general characteristics of the 78,356 Medicare decedents in our sample are as follows: the mean age is 81 years (range, 66–98), 56% are female; 40%, White; 36%, Black; and 11%, Hispanic. On average, women decedents are 3 years older than men decedents (mean age 82 vs. 79 years; $p < .001$) and have similar overall comorbidity (mean risk score for both genders is 2.2). In the 12 months before the final 6 months of life (the "pre-period"), more women receive Medicaid assistance (38% vs. 25%), and have more than 3 primary care visits (43% vs. 36%) than men (all p-values $< .001$; Table 1). The modal number of visits for both men and women is zero. Twenty-two percent of both genders have one or two visits, whereas only 18% of men have six or more visits, compared with 23% of women. In the final 6 months of life, both genders average 15 total hospital days (unadjusted); 74% of both men and women had at least one hospitalization. Compared with men, women cost $1,000 less per capita ($p < .05$) and use hospice a little more frequently (24% vs. 23%; $p < .001$) in the pre-period; they were more likely to die in a nursing home (29% vs. 20%; $p < .001$) and less likely in a hospital (41% vs. 45%; $p < .001$) or at home (19% vs. 22%; $p < .001$; Table 1).

Table 2 summarizes the unadjusted baseline comparisons by gender and nursing home resident status using each of our two definitions of nursing home residents: those with SNF bills from the MEDPAR files indicating a hospitalization during the pre-period, and those with BETOS code health care encounters from the carrier file received in a nursing home. We divided the population into four groups: Group 1 contains people who met both nursing home definitions (MedPAR and carrier file based); group 2 met the carrier file criterion only; group 3, MedPAR only; and, group 4, neither. During the pre-period, the

Table 1. Decedent Characteristics by Gender

Pre-Period Characteristic***	Total	Male	Female
n	78,356	34,302	44,054
Percent		44	56
Mean age (SD), yrs	80.9 (8.1)	79.1 (7.6)	82.3 (8.2)**
Race**			
White (%)	40	39	40
Black (%)	36	36	37
Hispanic (%)	11	11	10
Other (%)	14	14	14
Medicaid receipt (%)	32	25	38**
Mean comorbidity risk score‡ (SD)	2.2 (1.7)	2.2 (1.7)	2.2 (1.7)
Unadjusted end-of-life utilization†			
Mean total hospital days (SD)	15.1 (20.2)	15.3 (20.3)	15.0 (20.1)
Mean total costs, $1000 (SD)	24.9 (30.9)	25.4 (32.4)	24.4 (29.8)*
Any hospital admission (%)	74	74	74
Hospice (%)	24	23	24**
In-hospital death (%)	43	45	41**
Nursing home death (%)	25	20	29**
Death at residence (%)	21	22	19**

Abbreviation: SD, standard deviation.
* $p < .05$, difference between genders.
** $p < .001$, difference between genders.
*** Pre-period, months 18–7 before death.
† Measured during final 180 days of life.
‡ Comorbidity risk score determined by the relative risk score from DxCG's prospective risk adjustment software, which organized ICD-9-CM diagnosis codes from the Medicare utilization files, assigns weights to them, and summarizes their expected impact on future expenditures via a relative risk score.

comorbidity scores of the SNF groups 1 and 3 were higher than the non-SNF groups 2 and 4: namely, 3.6 for group 1, 2.7 for group 3, 1.5 for group 2, and 0.87 for group 4. During the pre-period, those with CARRIER file code definitions had more primary care visits than those who did not (41% of group 1 and 37% of group 2 had at least nine primary care visits, compared with 11% of group 3 and 7% of group 4); more Medicaid use (46% of group 1 and 55% of group 2, compared with 33% of group 3 and 28% of group 4); and less specialty use (14% for group 1, 7% for group 2 compared with 33% for group 3 and 46% for group 4). During the last 6 months of life, the carrier file only group 2 had the lowest costs ($10,000 compared with $17,000 for group 1, $19,000 for group 3, and $15,000 for group 4); lowest use of hospice (17%, compared with 28% of group 1, 29% of group 3, and 25% of group 4); and the most deaths in a nursing home (66%, compared with 55% of group 1, 33% of group 3, and 16% of group 4). Within each group, men received more specialty care, and more women had Medicaid and died in a nursing home.

When adjusting for all confounders except either of the two measures of nursing home use, we found that among those with nine or more primary care visits, women spent less time in the hospital than men (11.8 vs. 13.4 days; $p < .05$; Figure 1A), and averaged $3,677 less than men ($p < .05$; Figure 2A). After a threshold of three to five visits, more primary care visits were associated with fewer hospital days for both genders, but the association was stronger for women (Figure 1A). After a threshold of one to two visits, more primary care visits were generally associated with lower costs for women, but not for men (Figure 2A). Our results were similar when adjusting for nursing home use using MedPAR data: of those with nine or more primary care visits, women spent less time in the hospital than men (13.2 vs. 14.1 days, $p < .05$; Figure 1B) and averaged $2,641 less than men ($p < .05$; Figure 2B).

After a threshold of three to five visits, more primary care visits were generally associated with less hospital utilization and lower costs for women, but not for men (Figures 1B and 2B). However, when we adjusted for nursing home use using carrier file data (having at least two separated encounters in a nursing home defined by BETOS codes), we found that more primary care was not associated with decreased hospital utilization (Figure 1C) or costs (Figure 2C). Moreover, we did not find a significant gender difference in total hospital days, although women with at least nine primary care visits still had lower costs than men ($20,594 for women vs. $24,157 for men). When we stratified the population by gender and the carrier file definition of nursing home residence, both men and women residents had remarkably lower costs than nonresidents. After accounting for nursing home residence, there were no significant gender differences within these strata (Figures 1D and 2D).

Discussion

We sought to examine gender differences in health care utilization at the end of life. We expected nursing home use to be a confounder, because more women are nursing home residents than men, and such residence affects several kinds of Medicare utilization. Because Medicare billing data do not have a direct indicator of nursing home residence, we imputed two different proxy measures of nursing home use using variables from different Medicare file sources. When we did not control either at all, or adequately, for nursing home residence (using MedPAR files only), we reached conclusions different from those when using a more reliable marker of nursing home residence (based on receiving services, widely spaced in time, in a nursing home). This is an example of Simpson's paradox. Health care utilization outcomes are influenced by nursing home resident status, with costs and utilization being much lower for nursing home residents than nonresidents, and women receiving Medicare being far more likely than men to reside in nursing homes.

Table 2. Population of Nursing Home Residents, Characteristics by Gender and Two Definitions of Nursing Home Use

Pre-Period Characteristic*	Group 1 + SNF+ Carrier	Group 2 + Carrier- SNF	Group 3 + SNF- Carrier	Group 4 - SNF- Carrier
n	3949	8935	6089	59383
Men (%)	34	29	39	47
Women (%)	66	71	61	53
Medicaid (%)	46	55	33	28
Among men	42	51	27	21
Among women	48	56	36	34
Mean comorbidity score*** (SD)	3.6 (2.1)	1.5 (1.5)	2.8 (1.9)	0.88 (1.4)
Among men	3.9	1.8	3.0	0.90
Among women	3.4	1.4	2.7	0.87
Number of primary care visits, % in each category				
0	9	10	23	45
1-2	8	8	27	24
3-5	21	22	26	18
6-8	21	23	13	7
≥9	41	37	11	5
Mean % of specialty visits total physician visits (for those with any provider visits) (SD)	14 (25)	7 (20)	33 (35)	46 (40)
Among men	16 (27)	9 (21)	36 (36)	49 (40)
Among women	13 (24)	7 (19)	31 (35)	43 (40)
Unadjusted end-of-life utilization**				
Mean total costs, $1,000 (SD)	17 (22)	10 (16)	19 (24)	15 (20)
Hospice (%)	28	17	29	25
In-hospital death (%)	21	20	26	24
Nursing home death (%)	55	66	33	16
Among men	52	57	31	14
Among women	56	70	35	18

Abbreviations: BETOS, separated health care encounters in nursing facility based on Berenson-Eggers-Type-of-Service code 4B in the Medicare Carrier file; NH, nursing home; SD, standard deviation; SNF = nursing home use based on bills in Medicare MedPAR Skilled Nursing Facility file.

* Pre-period, months 18–7 before death.

** Measured during final 180 days of life.

*** Comorbidity risk score determined by the relative risk score from DxCG's prospective risk adjustment software, which organized ICD-9-CM diagnosis codes from the Medicare utilization files, assigns weights to them, and summarizes their expected impact on future expenditures via a relative risk score.

Our first method of imputing nursing home use (1 billed day in a SNF) from Medicare MedPAR files is frequently used when analyzing costs and utilization at the end-of-life Medicare billing data (Lubitz & Riley, 1993; Shugarman, Bird, Schuster, & Lynn, 2007). When we used this definition in our analyses, we found the association between more primary care visits and decreased hospital utilization and costs to be stronger for women than men. This finding is consistent with studies showing that, because women generally use more primary care than men, they may have more trusting relationships with their health providers, and more easily express their preferences for end-of-life care (Bertakis, Azari, Helms, Callahan, & Robbins, 2000; Bird et al., 2002). Compared with men, women may value a dignified death over receiving intensive, life-sustaining treatments (Bookwala et al., 2001). These findings suggest a gender-related disparity, in which men receive more expensive and intensive procedures than women at the end of life (Ayanian & Epstein, 1991; Bird et al., 2002; Valentin, Jordan, Lang, Hiesmayr, & Metnitz, 2003).

However, this way of identifying nursing home users only identifies beneficiaries who have been hospitalized and used Medicare-financed SNF care for a period of at most 90 days of rehabilitation, leaving a record in the Medicare's MedPAR (SNF) file. Beneficiaries in this group are heterogeneous, containing both outpatients who require nursing care or rehabilitation after a hospitalization and nursing home residents. It includes a mix of nursing home residents and nonresidents.

Our second definition of nursing home use (at least two BETOS codes in the Medicare carrier file for nursing home encounters, separated by at least 90 days) distinguished a population likely to be nursing home residents. This is confirmed by the high percentage Medicaid receipt, high numbers of primary care visits, high proportion of nursing home deaths, and lower overall costs. When we stratified the population with this definition of nursing home use, we found no gender differences among either nursing home residents or nonresidents. Thus, the gender differences we observed seem to be principally driven by the fact

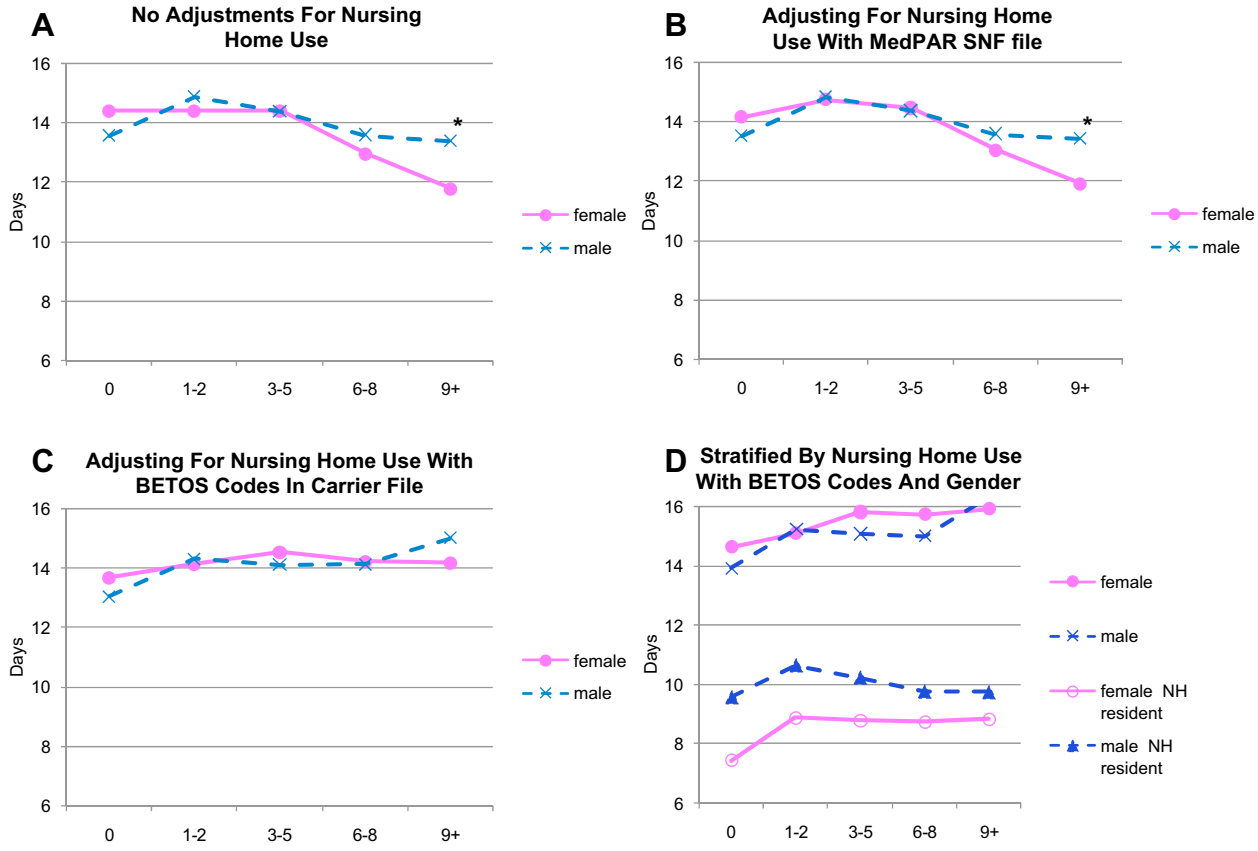


Figure 1. Preceding primary care visits** and total hospital days at the end of life***, by gender. **A,** No adjustments for nursing home use. **B,** Adjusting for nursing home use with MedPar SNF file. **C,** Adjusting for nursing home use with BETOS codes in carrier file. **D,** Stratified by nursing home use with BETOS codes and gender. *Abbreviations*: NH, nursing home; BETOS, separated health care encounters in nursing facility based on BETOS code 4B in the Medicare Carrier file; SNF = nursing home use based on bills in Medicare MedPAR Skilled Nursing Facility file. *$p < .05$, difference between genders. **Pre-period, months 18–7 before death; ***measured during final 180 days of life, adjusted for age, race, Medicaid use, comorbidity, geographic variation (hospital service area).

that far more women are nursing home residents, who (be they male or female) have lower hospital utilization and costs and more primary care visits than nonresidents. Our findings illustrate Simpson's paradox, in that relationships among utilization, cost, and primary care visits that seemed to differ by gender, disappear when beneficiaries were disaggregated by nursing home status. Although examples of Simpson's paradox have been described in clinical data (Charig, Webb, Payne, & Wickham, 1986) and meta-analyses (Cates, 2002), to our knowledge this is the first example of Simpson's paradox reported in Medicare data.

The difference between nursing home residents and nonresidents for both hospital utilization and costs was noteworthy. Nursing home residents had significantly more primary care visits than nonresidents, likely because Medicare nursing home regulations require at least one primary care visit every one or two months (*Medicare Claims Processing Manual*, N.D.). It is possible that regular primary care visits for nursing home residents may prevent hospitalization by better management of chronic conditions. In addition to using less specialty care, doctors who specialize in nursing home care may be better situated to implement directives that avoid hospitalization. Even when stratified by nursing home use, men received a higher proportion of their care by specialists compared with women.

The study has several limitations. Although we measured the quantity of primary care by counting visits to a primary care physician, we did not count other forms of primary care such as nurse visits, telephone consultations, or primary care provided by specialists. If there was a differential receipt of these services between men and women, these services could also be confounders when comparing genders in the receipt of primary care. Medicare claims data provide little to no information on the quality of primary care provided, the nature of clinician–patient interactions, beneficiary preferences, and the appropriateness of clinical treatment. Our findings do not address non-Medicare costs, such as for home health aides and nursing home residence, which can be considerable, and may not generalize to Medicare beneficiaries in managed care plans or to those without optional Medicare Part B (ambulatory care) coverage.

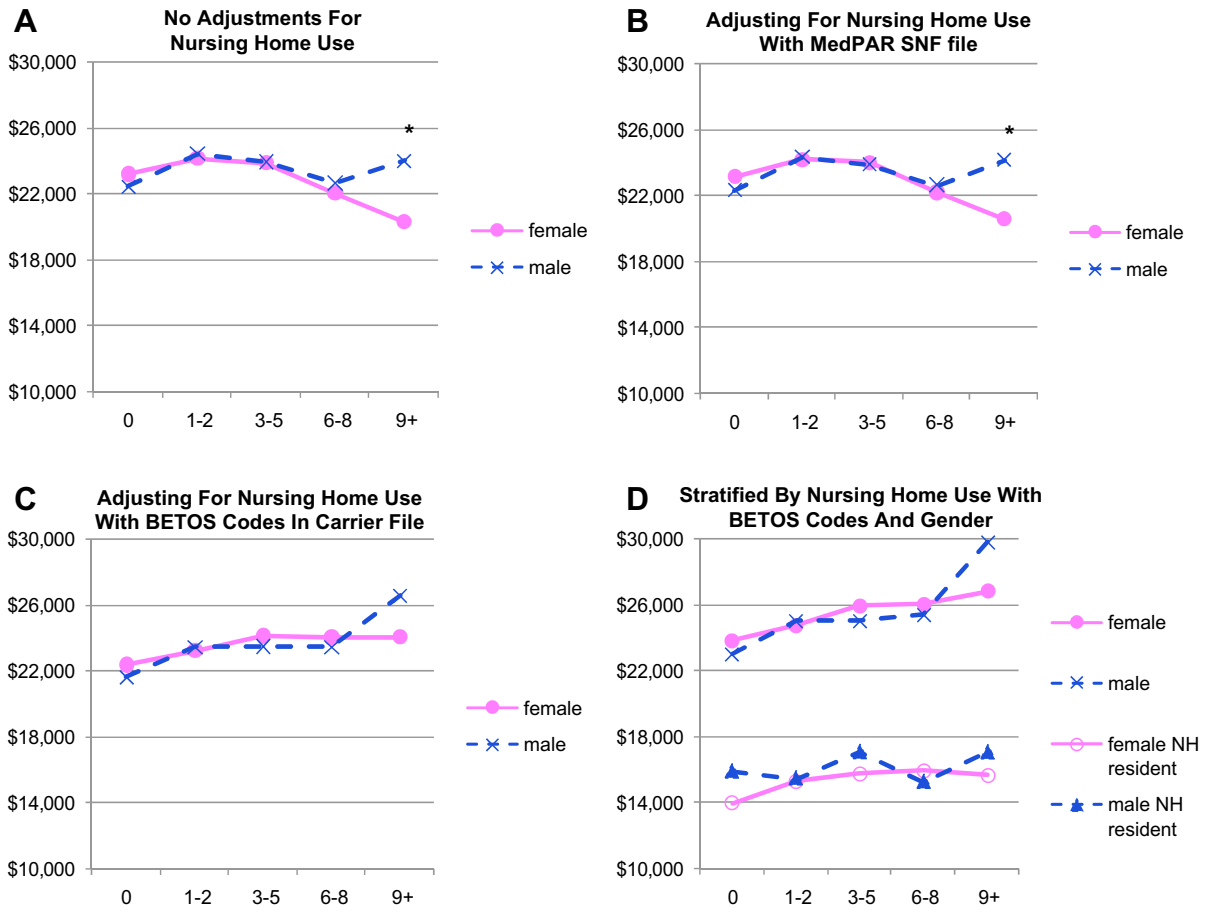


Figure 2. Preceding primary care visits‡ and ** total costs at the end of life***, by gender. **A**, No adjustments for nursing home use. **B**, Adjusting for nursing home use with MedPAR SNF file. **C**, Adjusting for nursing home use with BETOS codes in carrier file. **D**, Stratified by nursing home use with BETOS codes and gender. *Abbreviations*: NH, nursing home; BETOS, separated health care encounters in nursing facility based on Berenson-Eggers-Type-of-Service code 4B in the Medicare Carrier file; SNF = nursing home use based on bills in Medicare MedPAR Skilled Nursing Facility file. *$p < .05$, difference between genders; **Pre-period, months 18–7 before death; ***measured during final 180 days of life, adjusted for age, race, Medicaid use, comorbidity, geographic variation (hospital service area).

The decedent follow-back method that we used has been criticized for studying health care utilization before death (Bach, Schrag, & Begg, 2004). However, unlike Bach et al.'s study of cancer patients, 80% of our subjects did not die of cancer, meaning that there was likely not a well-defined moment at which such patients enter into a cohort with the expectation of a steady trajectory toward death (Lunney, Lynn, Foley, Lipson, & Guralnik, 2003). Most of our subjects have a vacillating trajectory of generally declining functional status resulting from chronic diseases such as heart failure and diabetes. So, although the prospective cohort approach may be more fruitful than the follow-back method for cancer patients, it is less clear how it might usefully be applied to a general population of elderly patients, with highly variable survival expectations.

In conclusion, when using Medicare data to examine gender-specific differences in health care utilization, it is important to account for nursing home use based on nursing home residence to avoid erroneous conclusions. Nursing home residence is highly confounded by gender because nursing home residents are predominantly female (Spillman, Liu, & McGilliard, 2002). We found that Medicare's MedPAR data does not identify nursing home residents, which can confound the interpretation of gender differences, even when studying only MedPAR (hospital or SNF) outcomes. However, we have shown how the potential confounding can be solved by using Medicare carrier file data, as described in this paper.

References

Approaching Death: Improving Care at the End of Life. (1997). Washington. DC: National Academy Press.

Ash, A. S., Ellis, R. P., Pope, G. C., Ayanian, J. Z., Bates, D. W., Burstin, H., et al. (2000). Using diagnoses to describe populations and predict costs. *Health Care Financing Review*, 21, 7–28.

Ash, A. S., Posner, M. A., Speckman, J., Franco, S., Yacht, A. C., & Bramwell, L. (2003). Using claims data to examine mortality trends following hospitalization for heart attack in Medicare. *Health Services Research*, 38, 1253–1262.

Ayanian, J., & Epstein, A. M. (1991). Differences in the use of procedures between women and men hospitalized for coronary heart disease. *New England Journal of Medicine*, 325, 221–225.

Bach, P. B., Schrag, D., & Begg, C. B. (2004). Resurrecting treatment histories of dead patients: A study design that should be laid to rest. *JAMA*, 292, 2765–2770.

Baicker, K., Chandra, A., Skinner, J. S., & Wennberg, J. E. (2004). Who you are and where you live: How race and geography affect the treatment of Medicare beneficiaries. *Health Affairs, Suppl Web Exclusive* VAR33–44.

Barnato, A. E., Lucas, F. L., Staiger, D., Wennberg, D. E., & Chandra, A. (2005). Hospital-level racial disparities in acute myocardial infarction treatment and outcomes. *Medical Care*, 43, 308–319.

Bertakis, K. D., Azari, R., Helms, L. J., Callahan, E. J., & Robbins, J. A. (2000). Gender differences in the utilization of health care services. *Journal of Family Practice*, 49, 147–152.

Bird, C. E., Shugarman, L. R., & Lynn, J. (2002). Age and gender differences in health care utilization and spending for Medicare beneficiaries in their last years of life. *Journal of Palliative Medicine*, 5, 705–712.

Bookwala, J., Coppola, K. M., Fagerlin, A., Ditto, P. H., Danks, J. H., & Smucker, W. D. (2001). Gender differences in older adults' preferences for life-sustaining medical treatments and end-of-life values. *Death Studies*, 25, 127–149.

Campbell, D. E., Lynn, J., Louis, T. A., Shugarman, L. R., Campbell, D. E., Lynn, J., et al. (2004). Medicare program expenditures associated with hospice use. *Annals of Internal Medicine*, 140, 269–277.

Cates, C. (2002). Simpson's paradox and calculation of number needed to treat from meta-analysis. *BMC Medical Research Methodology*, 2, 1.

Charig, C. R., Webb, D. R., Payne, S. R., & Wickham, J. E. (1986). Comparison of treatment of renal calculi by open surgery, percutaneous nephrolithotomy, and extracorporeal shockwave lithotripsy. *British Medical Journal Clinical Research Edition*, 292, 879–882.

Degenholtz, H. B., Thomas, S. B., & Miller, M. J. (2003). Race and the intensive care unit: Disparities and preferences for end-of-life care. *Critical Care Medicine*, 31(5 Suppl), S373–378.

DxCG. (2004). *DxCG risk adjustment software: User's guide release 6.1*. Boston: Author.

Edwards, C., & DeHaven, T. (2003). War between the generations: Federal spending on the elderly set to explode. *Policy Analysis*, 488, 1–22.

Fisher, E.S., & Wennberg, J.E. (2008). The Dartmouth atlas of healthcare. Available: http://www.dartmouthatlas.org/faq/data.shtm. Accessed February 2, 2008.

Fisher, E. S., Wennberg, J. E., Stukel, T. A., Skinner, J. S., Sharp, S. M., Freeman, J. L., et al. (2000). Associations among hospital capacity, utilization, and mortality of US Medicare beneficiaries, controlling for sociodemographic factors. *Health Services Research*, 34, 1351–1362.

Hanchate, A., Kronman, A. C., Young-Xu, Y., Ash, A. S., & Emanuel, E. (2009). Racial and ethnic differences in end-of-life costs: why do minorities cost more than whites? *Archives of Internal Medicine*, 169, 493–501.

Hogan, C., Lunney, J., Gabel, J., & Lynn, J. (2001). Medicare beneficiaries' costs of care in the last year of life. *Health Affairs*, 20, 188–195.

Institute of Medicine. (1997). Approaching death: Improving care at the end of life. Washington, DC: National Academy Press.

Johnston, J., & DiNardo, J. (1997). *Econometric methods* (4th ed.). New York: McGraw-Hill.

Kronman, A. C., Ash, A. S., Freund, K. M., Hanchate, A., & Emanuel, E. J. (2008). Can primary care visits reduce hospital utilization among Medicare beneficiaries at the end of life? *Journal of General Internal Medicine*, 23, 1330–1335.

Levinsky, N., Yu, W., Ash, A., Moskowitz, M., Gazelle, G., Saynina, O., et al. (2001). Influence of age on Medicare expenditures and medical care in the last year of life. *JAMA*, 286, 1349–1355.

Lubitz, J. D., & Riley, G. F. (1993). Trends in Medicare payments in the last year of life. *New England Journal of Medicine*, 328, 1092–1096.

Lunney, J. R. P. R. N., Lynn, J. M. D. M. A. M. S., Foley, D. J. M. S., Lipson, S. M. D., & Guralnik, J. M. M. D. P. (2003). Patterns of functional decline at the end of life. *JAMA*, 289, 2387–2392.

Medicare beneficiary demographics. (2006). Available: http://www.medpac.gov. Accessed July 27, 2008.

Medicare claims processing manual. (N.D.). Available: www.cms.hhs.gov. Accessed July 27, 2008.

Montgomery, J. E., Irish, J. T., Wilson, I. B., Chang, H., Li, A. C., Rogers, W. H., et al. (2004). Primary care experiences of Medicare beneficiaries, 1998 to 2000. *Journal of General Internal Medicine*, 19, 991–998.

Pritchard, R. S., Fisher, E. S., Teno, J. M., Sharp, S. M., Reding, D. J., Knaus, W. A., et al. (1998). Influence of patient preferences and local health system characteristics on the place of death. SUPPORT Investigators. Study to Understand Prognoses and Preferences for Risks and Outcomes of Treatment. *Journal of the American Geriatrics Society*, 46, 1242–1250.

Shugarman, L. R., Bird, C. E., Schuster, C. R., & Lynn, J. (2007). Age and gender differences in Medicare expenditures at the end of life for colorectal cancer decedents. *Journal of Women's Health*, 16, 214–227.

Skinner, J., Weinstein, J. N., Sporer, S. M., & Wennberg, J. E. (2003). Racial, ethnic, and geographic disparities in rates of knee arthroplasty among Medicare patients. *New England Journal of Medicine*, 349, 1350–1359.

Song, J., Chang, R. W., Manheim, L. M., & Dunlop, D. D. (2006). Gender differences across race/ethnicity in use of health care among Medicare-aged Americans. *Journal of Women's Health*, 15, 1205–1213.

Spillman, B.C., Liu, K., & McGilliard, C. (2002). Trends in residential long-term care: Use of nursing homes and assisted living and characteristics of facilities and residents. Available: http://aspe.hhs.gov/daltcp/Reports/rltct.htm. Accessed July 26, 2008.

Valentin, A., Jordan, B., Lang, T., Hiesmayr, M., & Metnitz, P. G. H. (2003). Gender-related differences in intensive care: A multiple-center cohort study of therapeutic interventions and outcome in critically ill patients. *Critical Care Medicine*, 31, 1901–1907.

Author Descriptions

Andrea C Kronman, MD, MSc, is an Assistant Professor of Medicine in the Women's Health Unit in the Section of General Internal Medicine at the Boston University School of Medicine. She is a clinician-investigator whose research experience includes using large databases to examine disparities in primary care and healthcare utilization at the end of life, as well as examining the relationship between depression and cancer screening behavior.

Amresh Hanchate, PhD, is a Health Economist at the VA Boston Healthcare System, and Research Assistant Professor in the Healthcare Research Unit in the Section of General Internal Medicine at the Boston University School of Medicine. He has many years of research experience in estimation of health costs, economic evaluation of health interventions, risk adjustment and health disparities.

Karen M. Freund, MD, MPH, is a Professor of Medicine, Chief of the Women's Health Unit, and Executive Co-Director of the Women's Health Interdisciplinary Research Center at the Boston University School of Medicine. She is a clinician-investigator whose research has focused on disparities in health outcomes for minority, low-income, and underserved women, focusing on cancer outcomes.

Ezekiel J. Emanuel, MD, PhD, is Head of the Department of Bioethics at The Clinical Center of the National Institutes of Health and a breast oncologist. He is currently acting as special advisor for health policy to the director of the White House Office of Management and Budget. He has published widely on the ethics of clinical research, health care reform, international research ethics, end-of-life care issues, euthanasia, the ethics of managed care, and the physician-patient relationship.

Arlene. S Ash, PhD, is Professor in the Department of Quantitative Health Sciences at the University of Massachusetts Medical School and Chief of its Division of Biostatistics and Health Services Research. She is adjunct Research Professor of Medicine and Chief of the Healthcare Research Unit at the Boston University School of Medicine. A health services research methodologist, she has studied health care delivery financing; patient safety; and racial and gender differences in medical utilization and health outcomes.

RESEARCH ARTICLE

AMERICAN JOURNAL OF medical genetics PART A

Consent for Genetic Research in the Framingham Heart Study

Daniel Levy,[1,2,3,4]* Greta Lee Splansky,[2] Nicolle K. Strand,[1,2] Larry D. Atwood,[2,3] Emelia J. Benjamin,[2,3] Susan Blease,[2] L. Adrienne Cupples,[2,6] Ralph B. D'Agostino Sr.,[2,7] Caroline S. Fox,[1,2] Margaret Kelly-Hayes,[2,3] Greg Koski,[5] Martin G. Larson,[2,7] Karen M. Mutalik,[2] Elizabeth Oberacker,[2] Christopher J. O'Donnell,[1,2] Patrice Sutherland,[2] Maureen Valentino,[2] Ramachandran S. Vasan,[2,3,7] Philip A. Wolf,[2,3] and Joanne M. Murabito[2,3]

[1]Center for Population Studies, National Heart, Lung, and Blood Institute, Bethesda, Maryland

[2]The National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, Massachusetts

[3]Departments of Medicine, Cardiology, Preventive Medicine and Neurology, Boston University School of Medicine, Boston, Massachusetts

[4]Cardiology Division, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts

[5]Departments of Medicine and Anesthesiology, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts

[6]Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts

[7]Department of Mathematics, Boston University, Boston, Massachusetts

Received 17 August 2009; Accepted 21 January 2010

Extensive efforts have been aimed at understanding the genetic underpinnings of complex diseases that affect humans. Numerous genome-wide association studies have assessed the association of genes with human disease, including the Framingham Heart Study (FHS), which genotyped 550,000 SNPs in 9,000 participants. The success of such efforts requires high rates of consent by participants, which is dependent on ethical oversight, communications, and trust between research participants and investigators. To study this we calculated percentages of participants who consented to collection of DNA and to various uses of their genetic information in two FHS cohorts between 2002 and 2009. The data included rates of consent for providing a DNA sample, creating an immortalized cell line, conducting research on various genetic conditions including those that might be considered sensitive, and for notifying participants of clinically significant genetic findings were above 95%. Only with regard to granting permission to share DNA or genetic findings with for-profit companies was the consent rate below 95%. We concluded that the FHS has maintained high rates of retention and consent for genetic research that has provided the scientific freedom to establish collaborations and address a broad range of research questions. We speculate that our high rates of consent have been achieved by establishing frequent and open communications with participants that highlight extensive oversight procedures. Our approach to maintaining high consent rates via ethical oversight of genetic research and communication with study participants is summarized in this report and should be of help to other studies engaged in similar types of research.

Key words: epidemiology; genetics; genome-wide association; medical ethics; population study

How to Cite this Article:
Levy D, Splansky GL, Strand NK, Atwood LD, Benjamin EJ, Blease S, Cupples LA, D'Agostino Sr. RB, Fox CS, Kelly-Hayes M, Koski G, Larson MG, Mutalik KM, Oberacker E, O'Donnell CJ, Sutherland P, Valentino M, Vasan RS, Wolf PA, Murabito JM. 2010. Consent for genetic research in the Framingham heart study.
Am J Med Genet Part A 152A:1250–1256.

Additional supporting information may be found in the online version of this article.

Grant sponsor: NIH; Grant number: N01-HC-25195.
This paper was conceived and drafted during Dr. Levy's Fellowship in the Medical Ethics Fellowship Program at Harvard Medical School.

*Correspondence to:
Daniel Levy, M.D., Framingham Heart Study, National Heart, Lung, & Blood Institute, 73 Mt. Wayte Avenue, Suite 2, Framingham, MA 01702.
E-mail: levyd@nih.gov

Published online 20 April 2010 in Wiley InterScience (www.interscience.wiley.com)
DOI 10.1002/ajmg.a.33377

INTRODUCTION

Since the beginning of the twentieth century, advances in the prevention and treatment of infectious diseases have led to a steady increase in childhood survival and in life expectancy. Today with people in developed and developing countries living longer, we have entered an era in which the greatest threats to global health are heart disease, cancer, stroke, and other adult chronic diseases. Most of these diseases are believed to be the result of interactions between genetic factors and environmental exposures. Extensive efforts are underway to understand the genetic underpinnings of complex diseases that affect the lifespan and quality of life of humans. Marked advances in technology, however, have ushered in new challenges to the appropriate use of genetic science to promote improvements in public health.

Genome-wide association methods have been applied selectively to individual diseases. Numerous genome-wide association studies of 100,000 to 1,000,000 or more single nucleotide polymorphisms (SNPs) are now underway in a range of sample sizes from a few hundred to up to tens of thousands of people to assess the association of common genetic variants with human diseases. Numerous novel genetic associations have recently been reported for scores of traits and diseases, including some that hitherto were resistant to genetic discoveries (http://www.genome.gov/gwastudies/).

The Framingham Heart Study is a prospective epidemiology project that began recruiting participants in 1948 [Dawber et al., 1951]. In recent years, study investigators have collected DNA samples and have prepared immortalized cell lines—to establish and maintain a renewable DNA resource—in study participants from three generations within families. Because of its wealth of data, multigenerational structure, and extensive DNA resources, the Framingham Heart Study is an attractive research setting for genome-wide association studies (GWAS). The National Heart, Lung, and Blood Institute initiated a GWAS in the Framingham Heart Study. This new project, the SNP Health Association Resource (SHARe), genotyped approximately 550,000 SNPs in over 9,000 study participants (approximately 5 billion genotypes) (http://public.nhlbi.nih.gov/GeneticsGenomics/home/share.aspx). This detailed characterization of common human genetic variation across the entire genome has helped pinpoint common genetic signatures of disease and thereby identified new pathways related to health and disease [Dehghan et al., 2008; Köttgen et al., 2009; Levy et al., 2009]. The success of this effort is dependent on high rates of consent by study participation for collection of biosamples and for the conduct of genetic research. High rates of consent are closely linked to the implementation of procedures for ethical oversight of genetic research, informed consent, access to data by outside investigators and for-profit companies, protection of privacy and confidentiality, and participant notification of genetic results. Left unaddressed, participants' concerns about the oversight of genetic research could impact rates of participation in the study. Accordingly, for this investigation, we calculated the rates of participant consent from 2002 to 2009 to various uses of their DNA and genetic information.

MATERIALS AND METHODS

Description of the Framingham Heart Study

In 1948, the Framingham Heart Study, under the auspices of the US Public Health Service, embarked on a prospective population-based study [Dawber et al., 1951]. A central objective of the study was to identify factors that contribute to cardiovascular disease by following its development over a long period of time in a large group of community residents who were extensively evaluated. The researchers recruited an original cohort of 5,209 men and women between the ages of 28 and 62, including 1,644 spouse pairs, from the town of Framingham, Massachusetts, and began the first round of physical examinations, lifestyle interviews and laboratory tests that they would later analyze for common patterns related to cardiovascular disease development [Dawber et al., 1951]. Since 1948, the original cohort participants have continued to return to the study every 2 years for a detailed medical history, physical examination, and laboratory tests. In 1971 the study, with the formal involvement of Boston University, began the enrollment of a second-generation cohort consisting of 3,548 children of the original cohort along with 1,576 of their spouses [Feinleib et al., 1975]. Offspring cohort examinations were similar to those of the original cohort and were repeated approximately every 4–8 years. Between 2002 and 2005, 4,095 adults with at least one parent in the offspring cohort enrolled in the third generation cohort and underwent a clinic examination [Splansky et al., 2007]. The second examination of that cohort began in 2008.

Consent for Genetic Research

At the start of the clinic visit, study participants provide written informed consent as part of a process that is administered by clinic staff trained to answer questions and seek a senior investigator to address participant questions they cannot answer. For the initial visit of the third generation cohort (2002–2005), separate check boxes were created to obtain consent (or refusal) for DNA extraction and sharing of DNA and genetic data with researchers, cell line creation, and access to DNA and data by for-profit companies (http://www.framinghamheartstudy.org/research/pdfs/consent/gen3_exam1_consent.pdf). The consent process and the consent document have evolved in response to ongoing discussion with ethicists and study participants. Additional check boxes were included in the offspring cohort Examination 8 consent document (2005–2008) to determine participants' preferences for the use of their data for specific research areas, including some that might be viewed as sensitive or not part of the historic core mission of the Framingham Heart Study (e.g., reproductive health, mental health and alcohol use) (http://www.framinghamheartstudy.org/research/pdfs/consent/exam8_offsite_consent.pdf). The third generation cohort's second examination consent form (in use since 2008) similarly includes a check box regarding the conduct of studies of potentially sensitive areas of research, as described above (http://www.framinghamheartstudy.org/research/pdfs/consent/gen3_exam2_consent.pdf). In both cohorts, permission was also explicitly obtained to notify participants (and with their permission, a designated personal physician) about the results of genetic tests that have important health and treatment implications. The notification

procedures are under development. Participant notification of genetic results will occur only when prespecified criteria are met: a genetic result has established analytic validity, the genetic variant poses significant and replicable risk for an important health condition, and proven therapeutic or preventive interventions exist for that condition [Bookman et al., 2006]. After initial consent was obtained to collect a cell line, the question was not repeated on subsequent consent forms. Similarly, when consent has been previously provided, questions are often removed from subsequent forms, in order to avoid unnecessary length and complexity.

For both cohorts, the right to withdraw from the study at any time is stated explicitly. It is important to note, however, that data sets have been created and distributed for public use (www.nhlbi.nih.gov/resources/deca/datasets_obv.htm). After they have been distributed we cannot go back and destroy data on participants who have withdrawn consent. We can only do so prospectively, by deleting their data from subsequent data sets. In addition, there has been continuous tracking of the level of permission for use of DNA samples and genetic research, especially protecting participants who have not been able to return for a recent examination. An annually updated database has been developed to track the most recent consent document provisions given by each participant as well as withdrawals of consent. The informed consent documents were reviewed and approved by the Institutional Review Board of Boston University Medical Campus. Based on our most recently updated consent information, we have recorded the consent preferences for each of the consent provisions at each examination and calculated the percentage of participants who consented to each provision.

RESULTS

Preferences About Participation in Genetic Research

Tables I–III summarize the number (and percent) of offspring and third generation cohort participants who granted or refused permission for each of multiple informed consent preference fields. This analysis is based on 2,980 offspring cohort participants, who attended their eighth clinic examination from 2005 to 2008, 4,095

TABLE I. Consent for Various Uses of DNA and Data in the Offspring Cohort (2005–2008)

Check box	Answer	Participants consented (%)
I agree to participate in the Framingham Heart Study examinations described above to study the frequency of and factors contributing to heart and blood vessel diseases, lung and blood diseases, stroke, memory loss, and other diseases and health conditions	Yes	2,980 (100)
	No	0 (0)
I agree to provide a blood sample from which DNA and other components can be extracted. The DNA will be made available to researchers studying the diseases listed above	Yes	2,891 (99.9)
	No	3 (0.1)
If a cell line has not already been collected, I agree to allow a cell line to be made from a sample of my blood to provide a renewable supply of DNA. (A cell line is a frozen sample of specially processed white cells from your blood that allows us to grow more white cells and get more DNA from them in the future as needed for research projects.)	Yes	2,969 (99.7)
	No	10 (0.3)
I agree to participate in the genetic studies of factors contributing to heart and blood vessel diseases, lung and blood diseases, stroke, and memory loss	Yes	2,978 (99.9)
	No	2 (0.1)
I agree to participate in genetic studies of other diseases and health conditions including but not limited to joint disease, bone loss, and cancer	Yes	2,974 (99.8)
	No	5 (0.2)
I agree to participate in genetic studies of reproductive conditions and mental health conditions such as alcohol use and depressive symptoms	Yes	2,970 (99.7)
	No	10 (0.3)
I agree to allow researchers from private companies to have access to my DNA and genetic data which may be used to develop diagnostic lab tests or pharmaceutical therapies that could benefit many people. (Note: You or your heirs will not benefit financially from this, nor will your DNA be sold to anyone.)	Yes	2,739 (91.9)
	No	240 (8.1)
If a genetic condition is identified that may have potentially important health and treatment implications for me, I agree to allow the Framingham Heart Study to notify me and with my permission to notify my physician	Yes	2,964 (99.5)
	No	16 (0.5)

Differences between the 2980 eligible individuals and the sum of responses for any check box reflect missing values.

TABLE II. Differences between the 4095 eligible individuals and the sum of responses for any check box reflect missing values.

Check box	Answer	Participants consented (%)
I agree to participate in the physical examination and genetic studies of factors contributing to heart and blood vessel diseases, lung and blood diseases, stroke, memory loss, joint disease, bone loss, deafness, cancer, and other major diseases and health conditions	Yes	4,095 (100)
	No	0 (0)
I agree to provide a blood sample from which DNA and other components can be extracted. The DNA will be made available to researchers studying the diseases listed above	Yes	4,092 (99.9)
	No	3 (0.1)
I agree to allow the creation of a cell line from my blood sample to provide a renewable supply of DNA. (A cell line is a frozen sample of specially processed white cells from your blood that allows us to grow more white cells and get more DNA from them in the future as needed for research projects.)	Yes	4,082 (99.7)
	No	12 (0.3)
I agree to allow researchers from private companies to have access to my DNA and genetic data which may be used to develop diagnostic lab tests or pharmaceutical therapies that could benefit many people. (Note: You or your heirs will not benefit financially from this, nor will your DNA be sold to anyone.)	Yes	4,000 (98.0)
	No	95 (2.0)

Differences between the 4095 eligible individuals and the sum of responses for any check box reflect missing values.

third generation cohort participants who attended their baseline clinic visit in 2002–2005, and 1,141 third generation cohort participants who attended their second clinic examination, which began in 2008 and is ongoing. Data from the two third generation cohort examination cycles were analyzed separately due to slightly different consent forms. More than 99% of participants attending the examination affirmatively selected check boxes for participation in the clinical examination and genetic studies, extraction of DNA and sharing of DNA with researchers, creation of a cell line for generating renewable DNA resource, use of genetic information for other purposes, including research that might be regarded as sensitive, and notification of genetic findings with health implications that might be discovered as a result of research. A total of 240 offspring participants and 17 third generation participants (8.1% and 1.5% respectively) did not permit sharing their DNA or genetic data with private companies. Data from the first generation

TABLE III. Consent for Various Uses of DNA and Data in the Third Generation Cohort Exam 2 (2008–2010)*

Check box	Answer	Participants consented (%)
I agree to participate in the FHS clinic examination and studies of the factors contributing to heart and blood vessel diseases, lung and blood diseases, stroke, memory loss, cancer, and other major diseases and health conditions	Yes	1,141 (100)
	No	0 (0)
I agree to provide a blood sample from which genetic material (DNA and other components) can be obtained. I agree to allow my data and blood samples to be used in the genetic studies of factors contributing to heart and blood vessel diseases, lung and blood diseases, stroke, memory loss, cancer, and other diseases and health conditions	Yes	1,140 (99.9)
	No	1 (0.1)
I agree to allow my data and blood samples to be used in genetic studies of reproductive conditions, and mental health conditions such as alcohol use and depressive symptoms	Yes	1,141 (100)
	No	0 (0)
I agree to allow researchers from commercial companies to have access to my DNA and genetic data which may be used to develop new lab tests or treatments that could benefit many people. (You or your heirs will not benefit financially from this, nor will your DNA be sold to anyone.)	Yes	1,123 (98.5)
	No	17 (1.5)
If a genetic condition is identified that may have important health and treatment implications for me, I agree to allow the FHS to notify me, and then with my permission to notify my physician	Yes	1,134 (100)
	No	0 (0)

Differences between the 1141 eligible individuals and the sum of responses for any check box reflect missing values.

[a]Exam cycle is still underway.

cohort was not obtained contemporaneously, so it was not analyzed for this study. However, they provided nearly universal approval at the last examination at which consent for DNA and genetic research were sought.

As of March 31, 2009 two individuals in the third generation cohort have withdrawn consent to participate in future clinic visits but maintained permission to use their previously collected data and DNA samples. Nine offspring cohort participants withdrew participation in further clinic examinations after examination cycle 7, which took place between 1998 and 2000. Two offspring participants withdrew consent to use their DNA. No offspring cohort participants have withdrawn after attending their eighth examination cycle (2005–2008).

DISCUSSION

This report found that more than 99% of offspring cohort participants at their eighth clinic examination and third generation cohort participants attending their first and second examination granted permission for DNA extraction and the creation of cell lines for genetic research. This is considerably higher than the 85% of participants in the National Health and Nutrition Examination Surveys who consented to genetic research in 2000 [McQuillan et al., 2003]. In telephone interviews of 489 randomly selected people from Pennsylvania, 25% said they would not be willing to participate in medical research and 29% indicated uncertainty about participation [Trauth et al., 2000]. The Multiethnic Study of Atherosclerosis recently reported its rates of consent for genetic research and, in that study, full consent was granted by 79% of participants [Green et al., 2006]. The higher rates of consent for genetic research in Framingham Heart Study participants may be due in part to the nearly 60 years legacy of the program and the family-based design of the offspring and third generation cohorts. Before they arrived in clinic for their baseline examinations, eligible participants were aware of the history of the Framingham Heart Study from local media coverage and from their family members. The study's focus on familial patterns of disease and the genetic aims of the study were described in recruitment materials. Thus, it is possible that eligible participants who strongly objected to genetic research declined study participation and their objections to consent for genetic research are not reflected in our data.

It is our belief, however, that our participants' high consent rates are also due in large part to our ongoing efforts to maintain communications with participants and to keep them informed about research activities and procedures including ethical oversight. We assert that the steps we have taken are vital to fostering the trust that is essential to maintain high rates of participation and retention in a prospective study. Beskow and Dean [2008] conducted a survey that confirmed our proposed explanation for the Framingham Heart Study's high rates of consent and retention. It was determined through communication with potential research subjects that when investigators take steps to protect participants' privacy and confidentiality and keep them updated and informed, participants develop increasing trust for the research institution; these steps in turn favorably affect rates of consent.

The sole area with more than nominal unwillingness of Framingham participants to grant consent was for sharing of DNA and genetic data with private companies. Active restrictions to private sector access to data or DNA by Framingham Heart Study offspring cohort participants occurred in 2000–2001 when considerable publicity about a for-profit company's attempt to sell Framingham Heart Study data resulted in participant concerns about such efforts [Ready, 2001]. That private venture did not move forward, in large part because of inconsistency with informed consent provisions.

It is clear from our close interactions with participants that they enrolled in the Framingham Heart Study and continue to attend periodic clinic visits out of a strong desire to contribute to a scientific effort to improve public health. Sharing data and DNA with the outside research community is critical to maximizing the scientific knowledge gained from participation in the study. There remains inherent tension, however, between measures to maximize sharing of data and DNA with the outside research community to promote scientific discovery and restrictive measures to protect participants' privacy and confidentiality. Acknowledging this balance, we have developed procedures to achieve both aims, but recognize that they will evolve over time. Several of these procedures are described below. For example, we distribute datasets free of charge via the National Heart, Lung, and Blood Institute (http://www.nhlbi.nih.gov/resources/deca/directry.htm), and distribute genetic and phenotypic data at no charge via dbGAP (http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000007.v6.p3). Simultaneously, we protect privacy by removing identifiers, requiring IRB approval, and executed data distribution agreements for investigators requesting DNA and databases.

Despite the belief of some scientists and ethicists that identifiable DNA should not be used for any research purpose other than that specifically stated in the consent document [McGuire and Gibbs, 2006], such restrictions of scientific use would have a chilling effect on discovery, because many questions that can be addressed in the long term using banked specimens are not apparent at the time of study inception. Recently, Kaye et al. [2009] described the importance of data sharing in genomics and the challenges that researchers face in maintaining the highest ethical standards and participant/donor privacy when they share their data with other investigators. They provided several recommendations including specific oversight of data sharing by a committee *other than* an Institutional Review Board and accurate and complete consent forms that cover all possible uses of DNA at recipient institutions without overwhelming participants. Framingham has been taking steps towards these ends for several years.

In 2003 a panel of medical ethicists convened by the Framingham Heart Study recommended that the study establish an ethics advisory board to make recommendations on ethics issues as they arise, and that, "the board include study participants as well as local clergy, physicians, genetic counselors and an ethicist." In response, we sent a newsletter to all study participants summarizing the panel recommendations and announcing plans to form an ethics advisory board with participant membership (http://www.framinghamheartstudy.org/participants/newsletters/spring2004.pdf). In early 2004 the Framingham Heart Study Executive Committee established the Framingham Ethics Advisory Board, chaired by a medical ethicist (GK), and comprised of a genetic counselor, two attorneys, two physicians and a clergyman from the community, as well as several Framingham Heart Study participants

representing each of the study cohorts. The Board has met approximately four times per year and the Framingham Heart Study has published its recommendations in newsletters sent to participants. Our approach is consistent with that of the Marshfield Clinic Personalized Medicine Research Project, which initiated conversations with participant focus groups, and formed an advisory board to improve communication and dialogue with study participants [McCarty et al., 2008].

The topic of large-scale genetics research studies, including a genome-wide association study, was discussed at several Ethics Advisory Board meetings. Participants have been regularly informed via newsletters about the rationale for and conduct of a number of genetic research projects. In November 2005, as more details about a potential genome-wide association study became known, the Ethics Advisory Board expressed its approval in concept for such a project and recommended convening a focus group of study participants to review the study aims and obtain feedback. Such a meeting was held in December 2005 and a list of general questions was generated by the study participants (see supporting information which may be found in the online version of this article). These were assembled with answers and included in the February 2006 newsletter to all study participants along with a general article about genome-wide association studies (http://www.framinghamheartstudy.org/participants/newsletters/winter2006.pdf). The questions raised by participants related to communicating study plans, protections of privacy and confidentiality, informing participants about genetic results, withdrawal of consent, sharing data with the scientific community, and commercial access to data and DNA. Timed to coincide with a national press release (http://www.nhlbi.nih.gov/new/press/06-02-06.htm), the February 2006 Framingham Heart Study newsletter also included a letter from the director of the National Heart, Lung, and Blood Institute describing plans to pursue a genome-wide association study in the three Framingham Heart Study cohorts in a manner consistent with participants' preferences. In addition, Framingham staff received educational session on the same topics.

Already in existence at this time was a system by which Framingham Heart Study participants' DNA and genetic or non-genetic data are distributed free of charge to outside investigators with several safeguards to protect privacy and confidentiality: (1) the investigator must submit and receive approval of a DNA application (http://www.framinghamheartstudy.org/univapp/index.php), which is reviewed by a DNA Committee composed of four members including a chairperson with no scientific relation to the Framingham Heart Study, (2) the investigator must obtain project approval from the applicant's Institutional Review Board and, (3) the investigator and host institution must execute a Data and Materials Distribution Agreement (http://www.framinghamheartstudy.org/research/proposal.html) with the National Heart, Lung, and Blood Institute and Boston University. The distribution agreement prohibits investigators from redistributing data or DNA to any third party and it prohibits any attempt to identify participants. After these three conditions are met, DNA and data are distributed to the investigator with a new and unique set of random identifiers. Although DNA for distribution is stripped of participants name and other identifying information (to prevent outright identification), it is not completely ''de-identified.'' In other words, it is theoretically possible for participants to be linked to their DNA, but attempting to do so violates the Distribution Agreement.

Similarly, once genome-wide association study planning was underway, we began to develop a set of procedures to provide broad data sharing while maintaining the security of participant data. To address these needs, we deposited our genotype and phenotype data in dbGaP, a secure online data-sharing repository that grants investigators access to genotype and phenotype data with various safeguards (http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study. cgi?study_id=phs000007.v1.p1). In order to protect the confidentiality and privacy of participant data that are stored in dbGaP a Data Access Committee (DAC) was formed. In order to access genetic and phenotypic data, investigators must submit a Data Use Certification (DUC) application, in which they agree to abide by dbGaP rules not to share participants' information with third parties or make any attempt to identify participants. This application is then reviewed by the DAC, which reserves the right to terminate access upon breach of dbGaP policy. Institutional Review Board approval from the investigators' institution is also required for access to Framingham dbGaP data.

From our discussions with Framingham Heart Study participants about genetic research, we have learned that key among their concerns are (a) the need for protections of privacy and confidentiality, (b) a desire that data and biological specimens be shared at no cost with the scientific community to maximize discoveries and improve public health, and (c) a need to honor restrictions of access to DNA and data by for-profit companies. These concerns, however, are likely to be universal and not exclusive to the Framingham Heart Study. Similar concerns about genetic research emerged in the deCODE genetics study in Iceland, including opposition by the Icelandic Medical Association [Annas, 2000]. Icelandic dissent centered on protections of privacy, for-profit use of DNA and data, lack of voluntary participation, restricted access to the data by the scientific community and inclusion of medical records in a for-profit database without specific individual consent. In addition, similar concerns were expressed by residents of British Columbia participating in a study of the public's attitudes towards informed consent in genetic research. Participants were most concerned with balancing their own confidentiality and privacy with ensuring that their DNA and data be available for useful research to promote the public good [Secko et al., 2009].

Whereas characterization of genetic variation across the human genome is now technically feasible and in widespread use, large-scale genetic studies must be carried out in a manner consistent with the preferences expressed by study participants in the informed consent process. This report describes the procedures implemented for ethical oversight of genetic research in the Framingham Heart Study, including the informed consent process, access to data by outside investigators and for-profit companies, and protections of privacy and confidentiality. We are engaged in ongoing communications with study participants to inform them about the goals of our genetic research program and seek their comments and concerns. Importantly, we established an Ethics Advisory Board that includes Framingham Heart Study participants to review genetic research and ensure its consistency with their consent and their wishes.

Moreover, given the myriad of questions that can be addressed via genetic studies, the Framingham Heart Study specifies multiple general areas of use in our current informed consent document (Tables I–III), while not specifying hundreds of potential areas of scientific inquiry. This practice educates our participants about the potential areas of research without overwhelming them with each and every conceivable possibility. Caulfield et al. [2008] recommended that whenever genome-wide association studies are conducted, participants should have the ability to withdraw consent at any time should they change their mind. The Framingham Heart Study has consistently notified participants of their right to withdraw consent for future distribution of any sample they have donated and has made a concerted effort to keep each participant's consent information up to date. The number of withdrawals of consent among Framingham Heart Study participants has been very low.

The Framingham Heart Study has succeeded in obtaining high rates of consent for genetic research and has taken multiple steps outlined above to implement ethical oversight of the large-scale genetic research that it is currently conducting. Lessons from the Framingham experience described in this report should be considered by other studies engaged in human subjects' research to assist in the development of procedures to ensure ethical oversight of genetic research in a manner consistent with participants' preferences. By developing detailed consent procedures and the ability to track decisions, means for communicating study plans with participants, and participant involvement in ethical oversight of genetic research, we have been able to maximize rates of participation in and consent for research, and we are hopeful that this pattern will be continued in the future.

ACKNOWLEDGMENTS

We would like to acknowledge the members of the Framingham Heart Study Ethics Advisory Board who volunteer their free time to make ours a better and more responsive study; the Framingham Heart Study participants who are the key to our past success and our future aspirations; and the Framingham Heart Study participant coordinators who administer our informed consent process. This study was supported by The Framingham Heart Study is funded by NIH (N01-HC-25195).

REFERENCES

Annas GJ. 2000. Rules for research on human genetic variation: Lessons from Iceland. N Engl J Med 342:1830–1833.

Beskow LM, Dean E. 2008. Informed consent for biorepositories: Assessing prospective participants' understanding and opinions. Cancer Epidemiol Biomarkers 17:1440–1451.

Bookman EB, Langehorne AA, Eckfeldt JH, Glass KC, Jarvik GP, Klag M, Koski G, Motulsky A, Wilfond B, Manolio TA, Fabsitz RR, Luepker1 RV. 2006. Reporting genetic results in research studies: Summary and recommendations of an NHLBI Working Group. Am J Med Genet Part A 140A:1033–1040.

Caulfield T, McGuire AL, Cho M, Buchanan JA, Burgess MM, Danilczyk U, Diaz CM, Fryer-Edwards K, Green SK, Hodosh MA, Juengst ET, Kaye J, Kedes L, Knoppers BM, Lemmens T, Meslin EM, Murphy J, Nussbaum RL, Otlowski M, Pullman D, Ray PN, Sugarman J, Timmons M. 2008. Research ethics recommendations for whole-genome research: Consensus statement. PLoS Biol 6:e73.

Dawber TR, Meadors GF, Moore FEJ. 1951. Epidemiological approaches to heart disease: The Framingham Study. Am J Public Health 41:279–286.

Dehghan A, Köttgen A, Yang Q, Hwang SJ, Kao WL, Rivadeneira F, Boerwinkle E, Levy D, Hofman A, Astor BC, Benjamin EJ, van Duijn CM, Witteman JC, Coresh J, Fox CS. 2008. Association of three genetic loci with uric acid concentration and risk of gout: A genome-wide association study. Lancet 372:1953–1961.

Feinleib M, Kannel WB, Garrison RJ, McNamara PM, Castelli WP. 1975. The Framingham offspring study. Design and preliminary data. Prev Med 4:518–525.

Green D, Cushman M, Desmond N, Johnson EA, Castro C, Arnett D, Hill J, Manolio TA. 2006. Obtaining informed consent for genetic studies. Am J Epidemiol 164:845–851.

Kaye J, Heeney C, Hawkins N, de Vries J, Boddington P. 2009. Data sharing in genomics-re-shaping scientific practice. Nat Rev Genet 10:331–335.

Köttgen A, Glazer NL, Dehghan A, Hwang SJ, Katz R, Li M, Yang Q, Gudnason V, Launer LJ, Harris TB, Smith AV, Arking DE, Astor BC, Boerwinkle E, Ehret GB, Ruczinski I, Scharpf RB, Ida Chen YD, de Boer IH, Haritunians T, Lumley T, Sarnak M, Siscovick D, Benjamin EJ, Levy D, Upadhyay A, Aulchenko YS, Hofman A, Rivadeneira F, Uitterlinden AG, van Duijn CM, Chasman DI, Paré G, Ridker PM, Kao WH, Witteman JC, Coresh J, Shlipak MG, Fox CS. 2009. Multiple loci associated with indices of renal function and chronic kidney disease. Nat Genet 41:712–717.

Levy D, Ehret GB, Rice K, Verwoert GC, Launer LJ, Dehghan A, Glazer NL, Morrison AC, Johnson AD, Aspelund T, Aulchenko Y, Lumley T, Köttgen A, Vasan RS, Rivadeneira F, Eiriksdottir G, Guo X, Arking DE, Mitchell GF, Mattace-Raso FU, Smith AV, Taylor K, Scharpf RB, Hwang SJ, Sijbrands EJ, Bis J, Harris TB, Ganesh SK, O'Donnell CJ, Hofman A, Rotter JI, Coresh J, Benjamin EJ, Uitterlinden AG, Heiss G, Fox CS, Witteman JC, Boerwinkle E, Wang TJ, Gudnason V, Larson MG, Chakravarti A, Psaty BM, vanDuijn CM. 2009. Genome-wide association study of blood pressure and hypertension. Nat Genet 41:677–687.

McCarty CA, Chapman-Stone D, Derfus T, Giampietro PF, Fost N. 2008. Community consultation and communication for a population-based DNA Biobank: The Marshfield clinic personalized medicine research project. Am J Med Genet Part A 146A:3026–3033.

McGuire AL, Gibbs RA. 2006. No longer de-identified. Science 312:370–371.

McQuillan GM, Porter KS, Agelli M, Kington R. 2003. Consent for genetic research in a general population: The NHANES experience. Genet Med 5:35–42.

Ready T. 2001. Framingham data not for sale. Nat Med 7:139.

Secko DM, Preto N, Niemeyer S, Burgess MM. 2009. Informed consent in Biobank research: A deliberative approach to the debate. Soc Sci Med 68:781–789.

Splansky GL, Corey D, Yang Q, Atwood LD, Cupples LA, Benjamin EJ, D'Agostino RB, Sr., 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.

Trauth JM, Musa D, Siminoff L, Jewell IK, Ricci E. 2000. Public attitudes regarding willingness to participate in medical research studies. J Health Policy 12:23–43.

American Pain Society
RESEARCH
EDUCATION
TREATMENT
ADVOCACY
PUBLISHED BY ELSEVIER

The Journal of Pain, Vol 11, No 11 (November), 2010: pp 1047-1055
Available online at www.sciencedirect.com

Clinical Factors Associated With Prescription Drug Use Disorder in Urban Primary Care Patients with Chronic Pain

Jane M. Liebschutz,*,† Richard Saitz,*,‡,§ Roger D. Weiss,¶ Tali Averbuch,# Sonia Schwartz,* Ellen C. Meltzer,* Elizabeth Claggett-Borne,* Howard Cabral,|| and Jeffrey H. Samet*,†

**Section of General Internal Medicine, Boston Medical Center, Clinical Addiction Research and Education (CARE) Unit, and Department of Medicine, Boston University School of Medicine, Boston, Massachusetts.*
†Department of Social and Behavioral Sciences, Boston University School of Public Health, Boston, Massachusetts.
‡Youth Alcohol Prevention Center, Boston University School of Public Health, Boston, Massachusetts.
§Department of Epidemiology, Boston University School of Public Health, Boston, Massachusetts.
||Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts.
¶Harvard Medical School and McLean Hospital, Boston, Massachusetts.
#University of Pennsylvania School of Nursing, Philadelphia, Pennsylvania.

Abstract: This study examined characteristics associated with prescription drug use disorder (PDUD) in primary-care patients with chronic pain from a cross-sectional survey conducted at an urban academically affiliated safety-net hospital. Participants were 18 to 60 years old, had pain for ≥3 months, took prescription or nonprescription analgesics, and spoke English. Measurements included the Composite International Diagnostic Interview (PDUD, other substance use disorders (SUD), Posttraumatic Stress Disorder [PTSD]); Graded Chronic Pain Scale, smoking status; family history of SUD; and time spent in jail. Of 597 patients (41% male, 61% black, mean age 46 years), 110 (18.4%) had PDUD of whom 99 (90%) had another SUD. In adjusted analyses, those with PDUD were more likely than those without any current or past SUD to report jail time (OR 5.1, 95% CI 2.8–9.3), family history of SUD (OR 3.4, 1.9-6), greater pain-related limitations (OR 3.8, 1.2-11.7), cigarette smoking (OR 3.6, 2–6.2), or to be white (OR 3.2, 1.7–6), male (OR 1.9, 1.1–3.5) or have PTSD (OR 1.9, 1.1–3.4). PDUD appears increased among those with easily identifiable characteristics. The challenge is to determine who, among those with risk factors, can avoid, with proper management, developing the increasingly common diagnosis of PDUD.

***Perspective:** This article examines risk factors for prescription drug use disorder (PDUD) among a sample of primary-care patients with chronic pain at an urban, academic, safety-net hospital. The findings may help clinicians identify those most at risk for developing PDUD when developing appropriate treatment plans.*

***Key words:** Primary care, substance abuse, pain.*

Chronic noncancer pain affects one-fifth of primary-care patients.[23] One pain treatment is opioid analgesic medication, which has been increasingly prescribed over the last decade[28,39,59] despite controversy about effectiveness.[20,36] Notwithstanding questions of efficacy, one risk to prescribing these medications is the potential of opioid addiction.[26,36,57]

The increase in prescribing has co-occurred with increasing misuse and abuse. This phenomenon ranges from misuse (nonmedical use, or for reasons other than prescribed) to the prescription drug use disorders (PDUDs) defined as abuse (misuse with consequences) or dependence (abuse with withdrawal or tolerance and/or uncontrolled use).[4] Rates of opioid misuse have increased since the 1990s, when lifetime misuse of pain relievers was less than 10%.[53] The 2007 National Survey

Received March 17, 2009; Revised August 11, 2009; Accepted October 27, 2009.
Supported by K23 DA016665 and K24 DA022288 from the National Institute of Drug Abuse and K24 AA015674 from the National Institute on Alcohol Abuse and Alcoholism.
Address reprint requests to Dr. Jane Liebschutz, Section of General Internal Medicine, Boston Medical Center, 801 Massachusetts Ave, 2nd Floor, Boston, MA 02118-2334. E-mail: jane.liebschutz@bmc.org
1526-5900/$36.00

doi:10.1016/j.jpain.2009.10.012

on Drug Use and Health (NSDUH) found that, among people over the age of 12, lifetime misuse of pain relievers increased to 13.3%. While 18.1% of the misused opioid analgesic medication came from legitimate physician prescriptions, 81% came from a friend or relative who was originally prescribed the medication by a physician.[16] Consequently, law enforcement efforts to stop opioid diversion have extended to physicians prescribing opioids for pain, adding fear of legal prosecution to provider concerns.[47] A further correlate of an increase in prescribing of opioids has been an increase in overdose, both intentional and unintentional.[21,40]

In this atmosphere, physicians report discomfort when prescribing opioid analgesics,[2,42,52] with particular concerns over the potential addiction risks and legal implications.[6,44,45,52] When patients have a known substance use disorder (SUD), physician discomfort about prescribing opioid analgesics increases, primarily due to suspicions that patient medication requests are motivated by addiction rather than pain.[37] Given these concerns about prescribing opioids, it would be useful to identify patient characteristics that are associated with prescription drug use disorder (PDUD).

Past research on the prevalence of and risk factors for PDUD among primary-care patients with chronic pain is limited by a number of factors: focus on a population referred for specialized pain treatment or who had been successfully maintained on more than 6 months of opioid therapy; diagnostic measures that have not been validated; or administrative databases. Ives et al[26] reported that 32% of patients referred to a primary-care-based chronic-pain disease management program were misusing opioids, similar to other studies.[10,43] Racial and ethnic minorities were a small fraction of patients studied.

Among patients treated with opioid analgesics risk factors for opioid misuse include SUDs,[26,34,38,43,56] family history of SUDs,[34,38] cigarette smoking,[10,38] legal problems,[26,38] younger age,[19,34,36,43] and higher doses of opioid medication.[38] Evidence of psychiatric comorbidity and higher pain severity were associated with opioid misuse in some studies,[19,34-36,38,56] but not others.[26]

Other studies have examined the risks for all substance use disorders for patients on opioids.[19] While national surveys have found that persons with PDUD often have overlapping polysubstance use disorders,[11,13,49,50] clinical samples have not examined how patients with PDUD may be similar or different from those with SUDs other than PDUD.

We therefore conducted a study to examine the clinical characteristics of individuals with lifetime PDUD, using validated measures among a population of primary-care patients with chronic pain drawn from an urban safety-net hospital setting. We also explored whether those with PDUD differ from those with SUDs other than PDUD. We hypothesized that factors similar to those identified in prior clinical studies of referral pain patients as well as epidemiological surveys would be present in an urban, minority sample. If clinicians are aware of relevant risk factors for PDUD in this population, they may be able to provide more informed clinical care.

Methods

Study Design

This was a cross-sectional study of primary-care patients with chronic pain, designed to examine characteristics of those with PDUD. Participants were recruited from waiting rooms of an academic, urban, safety-net[7] hospital primary-care practice that has 80,000 visits and 32,000 unique patients annually. Eligible patients were 18 to 60 years of age, spoke English, reported pain of 3 months or more, reported use of any analgesic medication (over-the-counter or prescription in the prior month), and had a scheduled primary-care appointment. Interviewers included both bachelor- and master-level research assistants, each trained for 2 weeks with ongoing quality control and principal-investigator supervision. Interviewers approached patients in primary-care waiting rooms per a predetermined pattern based on location in the waiting rooms and asked potential participants if they were interested in participating in a study on pain and health and to complete written screening questions. Eligible patients were told that the study interview would include questions about pain history, feeling sad or nervous, being hurt by someone, and using health care services. Informed consent was administered to eligible patients. Interviews lasted 45 to 90 minutes and participants were compensated $10. Recruitment occurred between February 2005 and August 2006. The Boston University Medical Center Institutional Review Board approved the study and Certificate of Confidentiality was obtained from the National Institutes of Health.

Participants

Of the 2,194 patients who answered the screening questions, 822 (37.4%) were eligible for the study, of whom 620 (75.4%) agreed to participate. Twenty-three participants did not complete enough of the interview due to time constraints including the SUD and PDUD assessment, and were excluded, leaving a sample of 597. When comparing those who enrolled with those who declined, enrollees were more likely to be black (61% vs 54.9%, $P = .04$), less likely to take over-the-counter pain medication (66.4% vs 78.5%, $P < .001$), and more likely to take opioid pain medication (41% vs 29.7%, $P = .002$).

Key Variables

Unless otherwise noted, all variables were obtained from subject interview.

Dependent Variables

Prescription Drug Use Disorder (PDUD) was defined as meeting DSM-IV criteria for lifetime sedative and/or opioid analgesic prescription drug abuse or dependence as measured by the Composite International Diagnostic Interview (CIDI) v.2.1 module on Drug Disorders.[4,58] Sedative was explicitly described, and included benzodiazepines and barbituates. Criteria for abuse included social, physical or legal consequences from use. The criteria for dependence additionally included compulsive use, health consequences, and physical dependence

(ie, tolerance or withdrawal). Physical dependence alone would not suffice to meet the diagnosis. Participants with PDUD could also have another SUD.

Other Substance Use Disorder (Other SUD) was defined as meeting DSM-IV criteria for any lifetime drug (excluding prescription) abuse or dependence measured using CIDI v 2.1 module on Drug Disorders, and/or past year alcohol dependence measured with the CIDI-short form (CIDI-SF) for alcohol dependence.[58] Current alcohol abuse and past alcohol use disorders were not measured as the CIDI-SF was used to reduce respondent time burden. Nicotine dependence was not included among other SUDs. We coded the mutually exclusive 3-level outcome variable as PDUD ± SUD, Other SUD (SUD alone), or no PDUD or SUD.

Independent Variables

We examined factors that we anticipated would be associated with PDUD or SUD, based on review of literature from clinical and epidemiological studies. We prioritized assessment of variables from a range of potential options based on strength of association with PDUD and SUD. For example, we chose to measure 2 major mental health disorders: depression and PTSD. Both are associated with pain, and PTSD is strongly associated with SUD. Other anxiety disorders were not measured because of the overlap with PTSD in a prior study in the same population.[33]A variety of specific violence measures were included to examine whether the independent associations with PDUD or SUD would be eliminated when PTSD or depression were considered. High pain severity and physical health-related quality of life have been suggested as predictors of PDUD,[34] as have prior incarceration,[26,38] cigarette smoking,[10,38] and family history of SUD.[38] The following independent variables were defined as: 1) sociodemographics including age (in years), gender, race/ethnicity (black, Hispanic, white, other), income (≥ or <$20,000), employment (unemployed or receiving disability payments vs other), education (< high school, high school+), marital status (partnered, divorced, single), health insurance (Medicaid/Medicare vs others, including private and uninsured); 2) pain-related limitations (high vs others) from the Graded Chronic Pain Scale, a 10-item validated measure of pain and disability;[54] 3) posttraumatic stress disorder (PTSD) diagnosis from the CIDI v. 2.1 PTSD module;[58] 4) major depression from the Patient Health Questionnaire (PHQ) for Depression, a 9-item validated measure correlated with past 2-week major depression;[30] 5) high adverse childhood experiences (ACE) (≥3 vs <3 experiences) as adapted from Felitti;[18] (Examples of ACE are physical abuse, neglect, sexual abuse, parental mental illness, single-parent household, parental incarceration, witnessing domestic violence); 6) intimate partner violence (IPV) —1 or more affirmative responses to questions adapted from the AddHealth Home Questionnaire;[1] 7) family history of SUD (single question about first-degree relatives with alcohol or drug problems); 8) jail time (single question about having spent time in jail); 11) SF-12 health-related quality-of-life physical- and mental-health composite scores;[55] 10) current cigarette smoking (taken from the Electronic Medical Record (EMR) visit closest to the interview date); and 11) opioid, benzodiazepine or sedative prescription in the past year (from the EMR).

Analysis

Bivariate analyses examined associations between 3 mutually exclusive groups: 1) lifetime PDUD (with or without other SUD); 2) lifetime other SUD (not including PDUD); and 3) no lifetime SUD or PDUD. If significant associations were found, then pair-wise comparisons were conducted. A multinomial, multivariable logistic regression model was then constructed predicting PDUD, other SUD and no lifetime SUD or PDUD, entering all independent variables found to be significant at the $P < .05$ level in the bivariate analyses. The results of this model are not shown here. We also included interaction terms which we suspected based on independent correlations and clinical intuition. We tested interactions between every possible pairing of variables within the 2 groupings: 1) Disability-related (3 pairs tested—insurance status, employment status, pain limitation); 2) Mental health/violence exposure (10 pairs tested—depression, SF-12 mental component summary score, PTSD, high-ACE, IPV-victimization. We then fit a subsequent parsimonious model which included all the independent predictors and the single significant interaction (IPV and depression). We also fit a separate logistic regression model with the same independent variables comparing only PDUD to other SUD. We then calculated the number and percentage of participants with each number of the variables found significant in the multinomial regression analysis in 3 mutually exclusive groups; PDUD, other SUD (ie, without PDUD), and no SUD or PDUD. We also calculated the sensitivity and specificity of number of factors for PDUD.

Results

Participant characteristics are shown in Table 1. Overall, a majority of the study sample was unemployed or disabled, poor, nonwhite, and experienced high pain disability. Forty percent had received an opioid prescription in the prior year. One hundred and ten (18.4%) met criteria for lifetime PDUD, of whom 99 (90%) had another SUD, and 61 (55.5%) met the PDUD diagnostic criteria in the past year. Fifty-five had both lifetime sedative- and opioid analgesic-use disorders, 44 had opioid analgesic-use disorder only, and 11 had sedative-use disorder only. Seventy-one percent of those with PDUD reported addiction to medications not prescribed for them. One hundred and forty-six (24.5%) had lifetime SUDs other than PDUD.

Bivariate Comparisons Between PDUD, Other SUDs and No History of SUDs

Participants with PDUD and other SUDs both had higher percentages of participants with the following characteristics compared to participants without SUD: male gender, Medicaid/Medicare, unemployment/

Table 1. Participant Characteristics, Stratified by Presence or Absence of Prescription Drug Use Disorder and Substance Use Disorder Diagnoses (n = 597)

					Pairwise Comparison		
Variable	Group 1 Prescription Drug Use Disorder (PDUD) N = 110 N(%)	Group 2 Other Substance Use Disorder* N = 146 N(%)	Group 3 No Substance Use Disorder N = 341 N(%)	Global P value	Group 1 vs. 2	Group 1 vs. 3	Group 2 vs. 3
Age in Years Mean (SD)	44.4 (9)	46.1 (8.1)	46.1 (10.4)	.2			
Sex							
Male	60 (54.5)	83 (56.9)	104 (30.5)	<.0001	.71	<.0001	<.0001
Female	50 (45.5)	63 (43.1)	237 (69.5)				
Race							
Black	51 (46.4)	90 (62)	222 (65.3)	<.0001	.02	<.0001	.12
Hispanic	10 (9.1)	14 (9.7)	35 (10.3)				
White	39 (35.4)	27 (18.6)	37 (10.9)				
Other	10 (9.1)	14 (9.7)	46 (13.5)				
Income							
<$20,000	62 (56.4)	89 (61)	214 (62.8)	.49			
≥$20,000	48 (43.6)	57 (39)	127 (37.2)				
Employment							
Employed	33 (30)	43 (29.5)	160 (46.9)	.0001	.92	.002	.0004
Unemployed	77 (70)	103 (70.5)	181 (53.1)				
Health Insurance							
Medicaid	80 (72.7)	100 (68.5)	183 (53.7)	.002	.83	.007	.007
Medicare	6 (5.5)	10 (6.9)	14 (4.1)				
Private Insurance	3 (2.7)	4 (2.7)	17 (5)				
Uncompensated Care State Fund	19 (17.3)	31 (21.2)	113 (33.1)				
Other	2 (1.8)	1 (.7)	14 (4.1)				
Education							
<High School	26 (23.6)	46 (31.5)	94 (27.6)	.38			
≥High School	84 (76.4)	100 (68.5)	247 (72.4)				
Marital Status							
Partnered	23 (20.9)	38 (26)	105 (30.8)	.28			
Separated/ Divorced/ Widowed	42 (38.2)	46 (31.5)	111 (32.5)				
Single	45 (40.9)	62 (42.5)	125 (36.7)				
Limiting Pain							
Yes	104 (94.5)	136 (93.1)	295 (86.5)	.02	.65	.03	.04
No	6 (5.5)	10 (6.9)	46 (13.5)				
Lifetime PTSD							
Yes	52 (47.3)	67 (45.9)	100 (29.3)	.0001	.82	.0006	.0005
No	58 (52.7)	79 (54.1)	241 (70.7)				
Current PTSD							
Yes	31 (28.2)	33 (22.6)	59 (17.3)	.04	.31	.01	.17
No	79 (71.8)	113 (77.4)	282 (82.7)				
Depression							
Yes	55 (50)	67 (45.9)	127 (37.2)	.03	.5	.02	.07
No	55 (50)	79 (54.1)	214 (62.8)				
≥3 ACE							
Yes	70 (63.6)	73 (50)	106 (31.1)	<.001	.03	<.0001	<.0001
No	40 (36.4)	73 (50)	235 (68.9)				
IPV Victim							
Yes	70 (63.6)	86 (58.9)	145 (42.5)	<.0001	.44	.0001	.001
No	40 (36.4)	60 (41.1)	196 (57.5)				
Family History of SUD							
Yes	76 (69.1)	96 (65.8)	114 (33.4)	<.0001	.57	<.0001	<.0001
No	34 (30.9)	50 (34.2)	227 (66.6)				
Time in Jail							
Yes	70 (63.6)	81 (55.5)	52 (15.3)	<.0001	.19	<.0001	<.0001
No	40 (36.4)	65 (44.5)	289 (84.7)				

Table 1. Continued

Variable	Group 1 Prescription Drug Use Disorder (PDUD) N = 110 N(%)	Group 2 Other Substance Use Disorder* N = 146 N(%)	Group 3 No Substance Use Disorder N = 341 N(%)	Global P value	Pairwise Comparison Group 1 vs. 2	Group 1 vs. 3	Group 2 vs. 3
Smoking							
Never	24 (23.3)	40 (28.6)	198 (61.5)	<.0001	.28	<.0001	<.0001
Previously	8 (7.8)	17 (12.1)	30 (9.3)				
Current	71 (68.9)	83 (59.3)	94 (29.2)				
Opioid Prescription in the Past Year							
Yes	47 (42.7)	59 (41.8)	132 (39.2)	.75			
No	63 (57.3)	82 (58.2)	205 (60.8)				
Benzodiazepine Prescription in the Past Year							
Yes	14 (12.7)	18 (12.8)	26 (7.7)	.13			
No	96 (87.3)	123 (87.2)	311 (92.3)				
Hypnotics Prescription in the Past Year							
Yes	7 (6.4)	7 (5)	17 (5)	.85			
No	103 (93.6)	134 (95)	320 (95)				
Pain Duration, years Mean (SD)	8.7 (8.6)	7.7 (8.3)	8 (9.1)	.6			
SF-12 Physical Health Mean (SD)	35.7 (11.1)	36.1 (11.5)	36.9 (11.9)	.6			
SF-12 Mental Health Mean (SD)	38.7 (11.5)	40.4 (12.3)	43.9 (12.9)	.0001	.03	.0002	.004

Abbreviations: PTSD, Posttraumatic Stress Disorder; ACE, High Adverse Childhood Experiences; IPV, Intimate Partner Violence; SUD, Substance Use Disorder; SF-12 Physical and Mental Health, Short Form-12 Physical and Mental Health Quality of Life.
*Other Substance Use Disorder includes participants with any substance use disorder, excluding prescription drug use disorder.

disability, severely limiting pain, PTSD, depression, intimate-partner victimization, family history of SUD, time spent in jail, and current smoking. Two measures showed differences between all 3 groups: mean SF-12 mental health component scores, and ≥3 ACEs. In both cases, the PDUD group had worse scores or greater exposures, respectively, compared to the other SUD group which in turn was worse than the no-SUD group. The PDUD group had a higher percent of whites compared to both other SUD and no SUD. The PDUD group had a higher percentage with current PTSD compared to those with no SUD but was not different compared to those with other SUD. No other factors differed significantly between the 3 groups (Table 1).

Independent Associations with PDUD and Other SUDs

In the first multinomial analysis, the following factors were found to be significantly associated with PDUD and/or other SUDs: time spent in jail, current smoking, family history of SUD, male gender, white race, higher limiting pain, PTSD, IPV, and depression. The interaction term between IPV and Depression was also statistically significant. In the final parsimonious model (which included only those 9 factors), time spent in jail, limiting pain, current smoking, family history of SUD, male gender, and PTSD were associated with a greater odds of both PDUD and SUD. White race was associated with PDUD. IPV and no depression, and No IPV with depression were associated with other SUDs (Table 2). In a logistic regression analysis comparing PDUD with other SUDs, the only significant association was white race (odds ratio 2.1, 95% CI 1.8–3.9) (other results not shown).

Table 3 shows the prevalence of PDUD and other SUD for each number of factors (0 through all 9). All patients with PDUD had 2 or more risk factors, and no person without SUD had more than 7 risk factors. The sensitivity and specificity of these combinations for PDUD are presented in Table 4.

Discussion

In the primary-care practice of an urban safety-net hospital, 18.4% of patients with chronic pain met criteria for a PDUD and an additional 24.5% had other SUDs. The vast majority of participants with PDUD had at least one co-occurring other SUD, and those with PDUD were virtually indistinguishable from those with other SUDs. Seven patient characteristics were independently associated with PDUD compared to those without any lifetime SUD: time spent in jail, high degree of

Table 2. Independent Associations With Prescription Drug Use Disorder and Other Substance Use Disorder Compared to Participants With no History of Substance Use Disorder* (n = 597)

Effect	PDUD Odds Ratio	95% CI	Other SUD Odds Ratio	95% CI
Time in Jail	5.11	2.8–9.34	3.82	2.22–6.57
Pain	3.77	1.21–11.72	2.76	1.1–6.91
Smoking	3.57	2.04–6.23	2.69	1.65–4.36
Family History Substance Abuse	3.40	1.93–6	3.41	2.08–5.6
White	3.17	1.68–6.01	1.48	.79–2.8
Male	1.94	1.07–3.53	2.37	1.39–4.03
PTSD	1.93	1.09–3.43	1.91	1.14–3.18
IPV and Depression	1.33	.62–2.87	1.12	.56–2.28
IPV and No Depression	1.92	.9–4.1	2.38	1.23–4.61
Depression and No IPV	1.84	.77–4.39	2.81	1.27–5.75

Abbreviations: PDUD, Prescription Drug Use Disorder; SUD, Substance Use Disorder; IPV, Intimate Partner Violence; PTSD, Posttraumatic Stress Disorder; CI, confidence interval.
NOTE. Reference group for IPV and depression interaction is the absence of IPV and depression.
*Multivariable multinomial logistic regression analyses, Pseudo R^2 = .42.

pain-related limitations, current smoking, family history of SUD, white race, male gender, and PTSD. The only difference between SUD and PDUD was the higher prevalence of white race in PDUD.

The finding that PDUD is highly associated with other SUDs corroborates prior clinical studies which relied on proxy measures for the determination of PDUD.[15,17,26,38,43,51] Ives et al[26] found that prior cocaine or alcohol abuse predicted prescription opioid misuse among patients at a pain management clinic. Studies which included smoking status as a correlate for PDUD found a striking association, as did our study. This reflects the exceptionally high comorbidity of smoking with alcohol and illicit-drug dependence found in other studies.[38,46]

Table 3. Prevalence of Prescription Drug Use Disorder (PDUD) and Other Substance Use Disorder Among Primary Care Patients by Number of Risk Factors* for PDUD and Other SUD (n = 597)

Number of Risk Factors*	Prescription Drug Use Disorder n = 110 N (%)	Other Substance Use Disorder n = 146 N (%)	No Substance Use Disorder n = 341 N (%)
0 (n = 9)	0 (0)	0 (0)	9 (100)
1 (n = 39)	0 (0)	2 (5.1)	37 (94.9)
2 (n = 94)	7 (7.5)	6 (6.4)	81 (86.2)
3 (n = 106)	10 (9.4)	19 (17.9)	77 (72.6)
4 (n = 125)	16 (12.8)	30 (24)	79 (63.2)
5 (n = 81)	14 (17.3)	32 (39.5)	35 (43.2)
6 (n = 80)	32 (40)	34 (42.5)	14 (17.5)
7 (n = 48)	24 (50)	15 (31.2)	9 (18.8)
8 (n = 10)	4 (40)	6 (60)	0 (0)
9 (n = 5)	3 (60)	2 (40)	0 (0)

*Risk Factors include: time spent in jail, family history of substance use disorder, smoking, white race, male gender, lifetime posttraumatic stress disorder, limiting pain, depression, intimate partner violence.

Table 4. Sensitivity and Specificity of Number of Risk Factors* for Prescription Drug Use Disorder (PDUD) (n = 451)

Number of Risk Factors*	Sensitivity %	Specificity %
> = 1 (n = 39)	100	<13.5
> = 2 (n = 94)	100	13.5
> = 3 (n = 106)	93.6	37.2
> = 4 (n = 125)	84.5	59.8
> = 5 (n = 81)	70	83.0
> = 6 (n = 80)	57	93.3
> = 7 (n = 48)	28	97.4
> = 8 (n = 15)	6	100

*Risk Factors include: time spent in jail, family history of substance use disorder, smoking, white race, male gender, lifetime posttraumatic stress disorder, limiting pain, depression, intimate partner violence.

For physicians who prescribe opioid and sedative medications, a critical concern is differentiating individuals who take medications for the intended therapeutic purposes from those with current, or potentially future, misuse and addiction. This cross-sectional study does not permit a longitudinal analysis of persons prescribed opioid analgesics. However, it does suggest that a propensity toward addiction (other SUD, family history of SUD, cigarette smoking) is a strong correlate of PDUD. Fleming et al[19] found that positive urine-toxicology screening for cocaine or marijuana and aberrant drug behaviors were among the significant predictors of SUD in primary-care patients receiving opioids. Data in our study suggest that those with PDUD have similar characteristics to those addicted to illicit drugs and alcohol. The 2005 report from the Center on Addiction and Substance Abuse at Columbia University corroborates this association in a community sample: 75% of persons who misuse prescription drugs have at least one co-occurring other SUD.[11] Of note, a physician may consider screening patients with pain and other risk factors for PDUD even in the absence of prescribing controlled substances, as most participants with PDUD in this study obtained the medication from sources other than a treating clinician.

A strong association of both PDUD and other SUDs was having spent time in jail. Almost two-thirds of those with PDUD reported having spent time in jail, compared to 15.3% of those with no SUD. The relationship between criminal activity and prescription drug abuse has been suggested in prior studies.[11,26] Akbik et al[3] reported prior legal problems predict subsequent opioid misuse among patients starting opioids for chronic pain. It is not known whether the jail history was due to crimes related to drug use, possession, manufacture or sale, which would suggest a history of SUD. It also may be a proxy for antisocial behavior, which is associated with PDUD.[24]

The associations of white race and male gender with PDUD in this sample reflect findings in other clinical and population samples.[8,13,24] Whites are prescribed more opioid medications in Emergency Departments and primary-care practices, perhaps reflecting a cultural bias by patients and physicians toward use of prescription opioids.[12,41] Male gender predominance reflects epidemiology of SUDs in general, although some data suggest more gender balance in PDUD. These trends suggest that future research should explore the social context, including gender and racial differences, of these associations.

Patients with PDUD reported a greater degree of pain-related limitation. Others have found low pain intolerance among those with active[14] and past addictions[32] compared to nonaddicted controls. It may be that lower pain threshold is an increased risk for developing addiction, or that addiction itself lowers the pain tolerance. This may complicate pain management among those with PDUD.

Among this sample of urban primary-care patients with chronic pain, PTSD was associated with PDUD. PTSD is known to be associated with SUD in clinical and community samples,[9,27,29,33] but the relationship with PDUD has not been described. The scientific evidence for neurological and physiological changes in PTSD,[25,48] pain,[5,6,31] and substance use disorders[22] is growing. Exploring these overlapping phenomena may allow development of tailored interventions.

Intimate partner violence and depression appeared to be individually associated with SUD but not PDUD, when accounting for the interaction between these phenomena. These associations may function differently in gender-specific analyses, which should be explored in future studies. The fact that Adverse Childhood Experiences were not independently associated with either PDUD or SUD after controlling for other variables suggests that its effect is mediated through other variables such as PTSD which remained significant in the final model.

These data strongly suggest that physicians treating patients with pain should assess for SUD. This can help direct care, including treatment for pain and substance-use disorders. Specialty pain practices who commonly prescribe opioid analgesics are likely to screen for this, but primary-care settings may not be as aware of the overlap between pain and addictions. Furthermore, patients do not always admit to SUD, particularly if they are intent on deceiving the treating physician to obtain prescription medication. Potential screening questions for patients with chronic pain could include assessment of smoking,and specific questions used in this study: "Do you have a family history of alcohol or drug problems?" and "Have you ever spent time in jail?" Evaluations for pain disability and PTSD may be additional clinical tools to help identify those at highest risk for PDUD.

After identifying a patient with risk factors, should clinicians prescribe opioid medications? In an observational study, Wiedemer et al[57] examined the impact of a structured opioid clinic for patients with risk factors, including psychiatric and substance use problems. All patients with SUD but no aberrant behaviors were safely maintained on opioids.[57] Clinical trials testing methods of opioid medication monitoring could inform clinicians about how to safely prescribe them to high-risk patients.

This study adds to the literature in 3 ways. First, it examines an urban, largely poor, and minority sample, which is underrepresented in the literature. Secondly, validated measures of PDUD used in this study improve upon the use of proxy measures that most other studies have employed. Finally, subjects were primary-care patients with chronic pain, not limited to those being prescribed chronic opioid medication. This can illuminate issues about patients who may require opioid analgesics in the future and common clinical concerns of urban primary-care pain patients.

The study limitations include possible misclassification of PDUD in individuals with pseudo addiction, ie, behaviors that resemble addiction but result from inadequate treatment of pain.[15] We believe that this is not a significant limitation, as the diagnostic criteria demand social or physical problems and compulsive use, which are not characteristic of pseudo addiction. Furthermore, 71% had addictions to medications that were not prescribed for them, which lowers the probability of misclassification. Another limitation was that lifetime alcohol-use disorders were not measured, which should attenuate any associations found because of misclassification bias. Since numerous independent predictors of PDUD and SUD were found, it is not clear how information on lifetime alcohol disorders would change the associations. The cross-sectional design limits conclusions regarding cause and effect; the findings would be strengthened by studying a longitudinal cohort. The recruitment strategy may limit the generalizability of the findings. However, as this was a study of PDUD risk factors and not prevalence, the associations should remain stable in a similar sample. Another limitation is that no corroborating evidence of prescription drug misuse was obtained, such as urine toxicology screening. Such testing can be helpful as a supplement to self-report.

In an urban cohort of primary-care patients with high levels of pain disability, unemployment and psychosocial stressors, PDUD was concentrated among those with: a family history of SUDs, having spent time in jail, current cigarette smoking, male gender, white race, pain-related functional limitations, and PTSD. The vast majority had co-occurring other SUDs. This suggests that clinicians could gain clinical insight by carefully evaluating such patients for these risk factors when developing a comprehensive pain-management strategy. It may also suggest which patients would benefit from a structured program for use of opioid medications. Refining the knowledge base on co-occurrence of addiction and pain could maximize safe and effective pain relief strategies.

Acknowledgments

The authors thank Denis Rybin, MSc for his assistance on data analysis. The uthors report no conflict of interest.

References

1. Add Health Questionnaire: Wave III In Home Questionnaire Code Book. The National Longitudinal Study of Adolescent Health

2. Adams N, Plane M, Fleming MF, Mundt M, Saunders L, Stauffacher E: Opioids and the treatment of chronic pain in a primary care sample. J Pain Symptom Manage 22:791-796, 2001

3. Akbik H, Butler SF, Budman SH, Fernandez K, Katz NP, Jamison RN: Validation and clinical application of the screener and opioid assessment for patients with pain (SOAPP). J Pain Symptom Manage 32:287-293, 2006

4. American Psychiatric Association, Diagnostic and Statistical Manual of Mental Disorders DSM-IV-TR, (4th ed.), 2000

5. Apkarian AV, Sosa Y, Sonty S, Levy RM, Harden RN, Parrish TB, Gitelman DR: Chronic back pain is associated with decreased prefrontal and thalamic gray matter density. J Neurosci 24:10410-10415, 2004

6. Baliki MN, Geha PY, Apkarian AV, Chialvo DR: Beyond feeling: chronic pain hurts the brain, disrupting the default-mode network dynamics. J Neurosci 28:1398-1403, 2008

7. Baxter RJ, Mechanic RE: The status of local health care safety nets. Health Aff (Millwood) 16:7-23, 1997

8. Blanco C, Alderson D, Ogburn E, Grant BF, Nunes EV, Hatzenbuehler ML, Hasin DS: Changes in the prevalence of non-medical prescription drug use and drug use disorders in the United States: 1991-1992 and 2001-2002. Drug Alcohol Depend 90:252-260, 2007

9. Bremner JD, Southwick SM, Darnell A, Charney DS: Chronic PTSD in Vietnam combat veterans: Course of illness and substance abuse. Am J Psychiatry 15:369-375, 1996

10. Butler SF, Budman SH, Fernandez K, Jamison RN: Validation of a screener and opioid assessment measure for patients with chronic pain. Pain 112:65-75, 2004

11. Califano JA: Under the counter: The diversion and abuse of controlled prescription drugs in the US. The National Center on Addiction and Substance Abuse at Columbia University. Available at: http://www.casacolumbia.org/absolutenm/articlefiles/380-Under%20the%20Counter%20-%20Diversion.pdf. Accessed September 1, 2007

12. Chen I, Kurz J, Pasanen M, Faselis C, Panda M, Staton LJ, O'Rorke J, Menon M, Genao I, Wood J, Mechaber AJ, Rosenberg E, Carey T, Calleson D, Cykert S: Racial differences in opioid use for chronic nonmalignant pain. J Gen Intern Med 2:593-598, 2005

13. Cicero TJ, Inciardi JA, Munoz A: Trends in abuse of Oxycontin and other opioid analgesics in the United States: 2002-2004. J Pain 6:662-672, 2005

14. Compton P, Charuvastra VC, Kintaudi K, Ling W: Pain responses in methadone-maintained opioid abusers. J Pain Symptom Manage 2:237-245, 2000

15. 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 16:355-363, 1998

16. Department of Health and Human Services: Results from the 2007 National Survey on Drug Use and Health: National Findings

17. Edlund MJ, Steffick D, Hudson T, Harris KM, Sullivan M: Risk factors for clinically recognized opioid abuse and dependence among veterans using opioids for chronic non-cancer pain. Pain 129:355-362, 2007

18. Felitti VJ, Anda RF, Nordenberg D, Williamson DF, Spitz AM, Edwards V, Koss MP, Marks JS: Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults. The Adverse Childhood Experiences (ACE) Study. Am J Prev Med 14:245-258, 1998

19. Fleming MF, Balousek SL, Klessig CL, Mundt MP, Brown DD: Substance use disorders in a primary care sample receiving daily opioid therapy. J Pain 8:573-582, 2007

20. Furlan AD, Sandoval JA, Mailis-Gagnon A, Tunks E: Opioids for chronic noncancer pain: A meta-analysis of effectiveness and side effects. CMAJ 174:1589-1594, 2006

21. 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 28(2):176-188, 2004

22. Goldstein RZ, Volkow ND: Drug addiction and its underlying neurobiological basis: Neuroimaging evidence for the involvement of the frontal cortex. Am J Psychiatry 159: 1642-1652, 2002

23. Gureje O, Simon G, Von Korff M: A cross-national study of the course of persistent pain in primary care. Pain 92: 195-200, 2001

24. Huang B, Dawson DA, Stinson FS, Hasin DS, Ruan WJ, Saha TD, Smith SM, Goldstein RB, Grant BF: Prevalence, correlates, and comorbidity of nonmedical prescription drug use and drug use disorders in the United States: Results of the National Epidemiologic Survey on Alcohol and Related Conditions. J Clin Psychiatry 67:1062-1073, 2006

25. Hull AM: Neuroimaging findings in post-traumatic stress disorder. Systematic review. Br J Psychiatry 181:102-110, 2002

26. Ives T, Chelminski P, Hammett-Stabler C, Malone R, Perhac J, Potisek N, Shilliday B, DeWalt D, Pignone M: Predictors of opioid misuse in patients with chronic pain: A prospective cohort study. BMC Health Services Research 6:46, 2006

27. Jacobsen LK, Southwick SM, Kosten TR: Substance use disorders in patients with posttraumatic stress disorder: A review of the literature. Am J Psychiatry 158:1184-1190, 2001

28. Joranson DE, Ryan KM, Gilson AM, Dahl JL: Trends in medical use and abuse of opioid analgesics. JAMA 283: 1710-1714, 2000

29. Kessler RC, Sonnega A, Bromet E, Hughes M, Nelson CB: Posttraumatic stress disorder in the National Comorbidity Survey. Arch Gen Psychiatry 52:1048-1060, 1995

30. Kroenke K: The PHQ-9: A new depression and diagnostic severity measure. Psychiatric Annals 32:509-521, 2002

31. Kuchinad A, Schweinhardt P, Seminowicz DA, Wood PB, Chizh BA, Bushnell MC: Accelerated brain gray matter loss in fibromyalgia patients: Premature aging of the brain? J Neurosci 27:4004-4007, 2007

32. Liebman PM, Lehofer M, Schonauer-Cejpek M, Legl T, Pernhaupt G, Moser M, Schauenstein K: Pain sensitivity in former opioid addicts. Lancet 344:1031-1032, 1994

33. Liebschutz J, Saitz R, Brower V, Keane TM, Lloyd-Travaglini C, Averbuch T, Samet JH: PTSD in urban primary care: high prevalence and low physician recognition. J Gen Intern Med 22:719-726, 2007

34. Manchikanti L, Cash K, Damron K, Manchukonda R, Pampati V, McManus C: Controlled substance abuse and illicit drug use in chronic pain patients: An evaluation of multiple variables. Pain Physician 9:215-226, 2006

35. Manchikanti L, Giordano J, Boswell M, Fellows B, Manchukonda R, Pampati V: Psychological factors as predictors of opioid abuse and illicit drug use in chronic pain patients. J Opioid Manag 3:89-100, 2007

36. Martell BA, O'Connor PG, Kerns RD, Becker WC, Morales KH, Kosten TR, Fiellin DA: Systematic review: Opioid treatment for chronic back pain: Prevalence, efficacy, and association with addiction. Ann Intern Med 146:116-127, 2007

37. Merrill JO, Rhodes LA, Deyo RA, Marlatt GA, Bradley KA: Mutual mistrust in the medical care of drug users: The keys to the "narc" cabinet. J Gen Intern Med17;327-333, 2002

38. 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 28:250-258, 2004

39. Olsen Y, Daumit GL: Opioid prescribing for chronic nonmalignant pain in primary care: Challenges and solutions. Adv Psychosom Med 25:138-150, 2004

40. Paulozzi LJ, Budnitz DS, Xi Y: Increasing deaths from opioid analgesics in the United States. Pharmacoepidemiol Drug Saf 15:618-627, 2006

41. Pletcher MJ, Kertesz SG, Kohn MA, Gonzales R: Trends in opioid prescribing by race/ethnicity for patients seeking care in US emergency departments. JAMA 299:70-78, 2008

42. Potter M, Schafer S, Gonzalez-Mendez E, Gjeltema K, Lobez A, QWu J, Pedrin R, Cozen M, Wilson R, Thom D, Croughan-Minihane M: Opioids for chronic nonmalignant pain. J Fam Pract 50:145-151, 2001

43. Reid MC, Engles-Horton LL, Weber MB, Kerns RD, Rogers EL, O'Connor PG: Use of opioid medications for chronic noncancer pain syndromes in primary care. J Gen Intern Med 17:173-179, 2002

44. Reidenberg MM, Willis O: Prosecution of physicians for prescribing opioids to patients. Clin Pharmacol Ther 81: 903-906, 2007

45. Richard J, Reidenberg MM: The risk of disciplinary action by state medical boards against physicians prescribing opioids. J Pain Symptom Manage 29:206-212, 2005

46. Richter KP, Gibson CA, Ahluwalia JS, Schmelzle KH: Tobacco use and quit attempts among methadone maintenance clients. Am J Public Health 91:296-299, 2001

47. Rosenberg T. When is a Pain Doctor a Drug Pusher? New York Times, June 17, 2007

48. Shin LM, Rauch SL, Pitman RK: Amygdala, medial prefrontal cortex, and hippocampal function in PTSD. Ann NY Acad Sci 1071:67-79, 2006

49. Simoni-Wastila L, Ritter G, Strickler G: Gender and other factors associated with the nonmedical use of abusable prescription drugs. Subst Use Misuse 39:1-23, 2004

50. Simoni-Wastila L, Strickler G: Risk factors associated with problem use of prescription drugs. Am J Public Health 94: 266-268, 2004

51. Sullivan MD, Edlund MJ, Steffick D, Unutzer J: Regular use of prescribed opioids: Association with common psychiatric disorders. Pain 119:95-103, 2005

52. Turk D: Clinicians' additudes about prolonged use of opioids and the issue of patient heterogeneity. J Pain Symptom Manage 11:218-230, 1996

53. US Department of Health and Human Services. Summary of Findings From the National Household Survey on Drug Abuse, 2007. Available at: http://www.oas.samhsa.gov/p0000016.htm#Standard. Accessed April 1, 2009

54. Von Korff M, Ormel J, Keefe FJ, Dworkin SF: Grading the severity of chronic pain. Pain 50:133-149, 1992

55. Ware J Jr., Kosinski M, Keller SD: A 12-Item Short-Form Health Survey: Construction of scales and preliminary tests of reliability and validity. Medical Care 34:220-233, 1996

56. Webster LR, Webster RM: Predicting aberrant behaviors in opioid-treated patients: Preliminary validation of the opioid risk tool. Pain Med 6:432-442, 2007

57. Wiedemer NL, Harden PS, Arndt IO, Gallagher RM: The opioid renewal clinic: A primary care, managed approach to opioid therapy in chronic pain patients at risk for substance abuse. Pain Med 8:573-584, 2007

58. Robins LN, Wing J, Wittchen H-U, Helzer JE, Babor TF, Burke J, Farmer A, Jablensky A, Pickens R, Regier DA, Sartorius N, Towle LH. The Composite International Diagnostic Interview: An epidemiologic instrument suitable for use in conjunction with different diagnostic systems and in different cultures. Arch Gen Psychiatry 45:1069-1077, 1988

59. 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. Medical Care 44: 1005-1010, 2006

A Chasm Between Injury and Care: Experiences of Black Male Victims of Violence

Jane Liebschutz, MD, MPH, Sonia Schwartz, BA, Joel Hoyte, BA, Lauren Conoscenti, PhD, Anthony B. Christian, Sr., Leroy Muhammad, Derrick Harper, MSW, LCSW, and Thea James, MD, MPH

Background: Despite higher rates of stabbing and shooting violence among black men, healthcare systems have not demonstrated an efficacious response to these patients. This study describes challenges and promotive factors for engaging black male violence victims of violence with medical and mental healthcare.

Methods: Black male victims of stabbings and shootings were recruited through fliers and word of mouth, and were interviewed individually (n = 12) or in pairs (n = 4) using a semistructured guide. A racially diverse multidisciplinary team analyzed the data using Grounded Theory methods.

Results: Challenges to engagement with healthcare included the following: (1) Disconnect in the aftermath; e.g. participants reported not realizing they were seriously injured ("just a scratch" "poke"), were disoriented ("did not know where I was"), or were consumed with anger. (2) Institutional mistrust: blurred lines between healthcare and police, money-motivated care. (3) Foreshortened future: expectations they would die young. (4) Self-reliance: fix mental and substance abuse issues on their own. (5) Logistical issues: postinjury mental health symptoms, disability, and safety concerns created structural barriers to recovery and engagement with healthcare. Promotive factors included the following: (1) desire professionalism, open personality, and shared experience from clinicians; (2) turning points: injury or birth of a child serve as a "wake up call"; and (3) positive people, future-oriented friends and family.

Conclusions: For black male violence victims, medical treatment did not address circumstances of and reactions to injury. Policies delineating boundaries between medical care and law enforcement and addressing postinjury mental health symptoms, disability, and safety concerns may improve the recovery process.

Key Words: Black male, Community violence, Qualitative research.

(*J Trauma.* 2010;69: 1372–1378)

Hand gun and stabbing violence disproportionately affects young African American men in the United States.[1,2] Although homicides are the most extreme health effect of violence, injuries, with their attendant physical and psychologic morbidity, greatly impact survivors.[3,4]

Despite the burden of violence on young black males, limited literature describes effective healthcare responses to shooting and stabbing violence. An opportunity exists to engage those injured from violence while receiving care in acute hospital settings. Existing hospital-based programs focus on case management approaches when working with violence survivors, often referring them for community-based services,[5–8] such as education and vocational help.[9] These programs showed reduced criminal activity and re-injury compared with patients receiving care as usual.[6–8]

Although the literature specifically examining the effectiveness of healthcare response to black male victims of violence is limited, there is an extensive literature outlining disparities in healthcare treatment among different racial and ethnic groups. For instance, an Institute of Medicine 2002 report suggested that aspects of the clinical encounter, patient preferences, physician stereotyping, and system structure may contribute to these disparities.[10] Also, in a study to understand perceived barriers to and preferences for general healthcare among black men, Ravenell et al.[11] found that trauma survivors cited cultural differences, as well as a lack of awareness of medical concerns, as barriers to care. They also noted fear of serious conditions and fatalism ("I'm not going to make it anyway") as reasons to avoid medical care.

We performed a qualitative study using Grounded Theory methods to understand the experiences of survivors of gun shot and stab wounds, to gain insight for planning interventions in hospitals that treat victims of community violence. The data from the following study describes the potential challenges to and promotive factors for, engaging black male victims of gun shot and stabbing injuries with medical and mental healthcare.

MATERIALS AND METHODS

Study Design

This was a cross-sectional study of an urban, community-based sample of black male victims of gunshot or stabbing injuries. Qualitative data were collected through semistructured interviews and analyzed using a Grounded Theory approach.[12] This study obtained approval from the Institutional Review Board and a Certificate of Confidentiality from the National Institutes of Health.

Submitted for publication January 29, 2010.
Accepted for publication May 11, 2010.

From the Clinical Addictions Research and Education Unit (J.L., S.S., J.H.), Section of General Internal Medicine, and Department of Emergency Medicine (T.J.), Boston Medical Center, Boston University Schools of Medicine and Public Health; National Center for PTSD (L.C.), VA Boston Healthcare System; Violence Intervention Advocacy Program (A.B.C., L.M.), Boston Medical Center; and Pyramid Builders Association (D.H.), Boston, Massachusetts.
Address for reprints: Jane Liebschutz, MD, MPH, Boston Medical Center, 801 Massachusetts Avenue, 2nd Floor, Boston, MA 02118; email: Jane.Liebschutz@bmc.org.

DOI: 10.1097/TA.0b013e3181e74fcf

Selection of Participants

Participants were recruited via fliers posted in the community and through word of mouth. Eligibility criteria included a history of emergency medical treatment for a stabbing or shooting injury, aged 18 years to 38 years, male gender, and self-identified black race. Eligible participants provided written informed consent before an interview and received $25 compensation following the interview.

Procedures

Trained interviewers conducted 14 semistructured interviews (12 individual interviews and 2 interviews with 2 participants each) from January 2008 to December 2008. The interviewers were two community-based, black, male, mental health professionals (seven interviews), a white female research assistant (three interviews), a black male research assistant (three interviews), and the principal investigator (one interview). The digitally recorded interviews lasted 45 minutes to 120 minutes. The interview guide was developed to include open-ended questions about the injury, experiences with healthcare after the injury, encounters with primary care, and views on mental health treatment, substance use, and research. Interviewers also inquired about experiences with race and gender if not spontaneously mentioned by participants. Audio-recordings were professionally transcribed and all identifying information was removed to protect the participants. The research team reviewed the interviews and then updated the interview guide so that future participants could clarify concepts found in earlier interviews. Participants also filled out a survey including demographics, posttraumatic stress symptoms (Posttraumatic Checklist-Civilian),[13–15] and alcohol use (AUDIT).[16]

Primary Data Analysis

The primary data analysis was conducted by an interdisciplinary and racially diverse research team consisting of two physicians (Emergency Medicine, Internal Medicine), two research assistants, a clinical psychologist, a mental health counselor, and two patient advocates working with victims of violence. Three members of the team were white and five were black. Additionally, one member of the team had a history of gun shot injury. The study used Grounded Theory[12] to analyze semi-structured interviews. Transcripts were read and audio-recordings listened to multiple times by the research team to identify common words and phrases of similar meaning to classify and create concepts. Researchers developed a coding scheme through discussion of each concept identified and placed excerpts from the transcripts into the codes with the aid of N-VIVO version 7 software.[17] At least two authors coded each transcript, and the research team discussion resolved coding discrepancies. The entire research team together grouped codes to create larger conceptual themes. For the final analysis, the research team generated theoretical constructs by examining these themes in the context of published literature from public health, medicine, social science, and psychology. The clinicians and advocates had extensive experience treating patients with stabbing and shooting injuries, which helped inform their interpretations of the themes. The findings were then placed into the context of published literature for the final analysis and presentation.

RESULTS

Characteristics of Study Subjects

Of the 16 participants enrolled in the study, the median age was 31 years (range, 25–38 years). Three were victims of gunshot, five of stab wounds, and eight of both injuries. The median length of time since the most recent injury was 5.5 years (range, 4 months–20.1 years). The median Posttraumatic Checklist-Civilian score was 43 (range, 22–69) and the median AUDIT score was 9 (range, 0–32).

Main Results

Researchers were struck by a lack of descriptions of interactions with healthcare clinicians, even when asked specifically about care after injury. For example, one participant described in exquisite detail all aspects of a stabbing assault. He summed up his medical treatment as, "let me see,… they took me to the emergency room. After that, for follow-up appointments…" This lack of detail in the narratives may indicate a sense of disconnection from the healthcare setting experienced by these men. Several themes, described in detail below, may elucidate the potential causes of this divide, and potential facilitators for future engagement with healthcare personnel. Table 1 provides illustrative quotations of the main themes.

Barriers to Care

Disconnect

Perceptions of the injury may be the source of a lack of participation and cooperation with clinicians. Eight participants reported not realizing they were injured in the moment of trauma or the extent of the injury. The language to describe the injury may also typify this disconnect: "scratch" (a liver laceration and a gunshot wound to the arm) and "poked" (multiple stabbings in abdomen).

Participants may not necessarily perceive the life-threatening aspect of injuries in the immediate aftermath. Their experience of the injury and the foreign environment of the hospital may lead to the development of a different injury awareness, which when considering the severity of the clinical observations, may cause both patients and clinicians to misinterpret each other. As described by one participant, "I know it's a real serious injury, but I saw I was alive, and ,. . . I didn't think it was a big deal at that time, 'cause I'm like, "I'm alive!"… And y'all over here, actin' like. . . it was real serious."

A range of emotions flooded participants once they realized what occurred. All participants described an overwhelming sense of anger in the immediate aftermath of the injury. For many, thoughts of anger, and possible retaliation, were all consuming, "But, when you first get stabbed, your mind goes blank. You don't think about anything but just revenge." One consequence of the anger was a desire to leave the hospital against medical advice to deal with their feelings. "[I was in the hospital] only three days, 'cause I was so mad that I wanted to leave." Furthermore, waking up from emergent surgery or lost consciousness, experienced by eight participants, contributed to the heightened emotional state by creating confusion and surprise.

TABLE 1. Barriers to and Promotive Factors for Engaging Black Male Community Violence Victims With Health Care

Theme	Subconcepts	Participant Quote
Physical and emotional disconnect	Not realizing that they had been injured	"I seen my cousin fall first, and I was askin' him, 'Are you a'right? And then I just noticed that he was bleedin, I was like, 'Yo, you know you're bleedin'?' He was like, 'No, I think that's you.' I checked myself, I didn't see nothin,' so I said, 'No, it must be you,' and it was . . . he got shot twice in his legs. I got up and I felt somethin' warm on my side, and he was like, 'Yo I think you (got) shot too,' and I looked and I was, I was hit in the side."
	Confusion after waking up	"When I first woke up out the coma, I didn't know what the f**k was goin' on, I didn't even know WHERE I was at. I woke up, like, 'Damn!' I'm lookin' around . . . can't even talk . . . and then I looked down at my stomach . . . 'Oh, what the f**k???' Now my heart's racin,' like, 'Damn!' I'm lookin' at my insides," and like, my f**kin' stomach's this wide . . . open!"
	Anger	"When I was in the hospital, there was nothin' nobody could tell me, like, "Naw, like f**k this. Like, I'm through with it. Like, I'm gonna kill everybody . . . like, anybody who have problems with me, I was gonna deal with, or if anybody looked at me wrong, I was gonna have . . . I was gonna do somethin' to."
Institutional mistrust	Police	"Well, the cops, as usual, were a**holes. Like it was our fault we got shot. First question they asked is, "Do you know who did it?" And then, "What did you do to the person that did it to you? Whose your enemies?" . . . they was actin' like you just don't get shot, it couldn't be an accident, that we was walkin' to the block and they got the wrong person; or they were lookin' for somebody else. And that was the case! The dude who shot us, it had nuttin' to do with us."
	Suspicion of clinician motives	"But they were happy to have people of color to patch up and profit off of. So, I guess they were grateful for that way. But, other than that, I was treated indifferently, especially by the nurses and the, um, after-care portion, the few days I spent in the hospital, I might as well have been a piece of meat."
Foreshortened future	Expect to die young	"When I was growin' up, lot of cats and neighborhood people used to say most kids wasn't gonna make it to see twenty-one."
Self-reliance	Mental health	"And, especially from the 'hood, if that's what you want to call the inner city, they think that the counselors don't care. They don't care about what's really goin' on in my life, or how I feel . . . Like, they're gettin' paid to sit here and play with my life like a yo-yo, and then people feel like, 'Oh, they don't care.'"
	Substance abuse	"Well, I cope with [PTSD symptoms] . . . through the grace of god, beer, and weed . . . (laughs). Well, that's not the right solution, but I told you, I ain't had no help, no counselin' or nothin' like that."
Preferred clinician characteristics	Race	"Like I said, I never had no problems with the doctors. It was just like they was . . . they was basically there for me It was just like . . . to me, it was a surprise, and like, different people, like . . . different colors, different race, you come through, they really don't care about you, but . . . these people showed me they're sorta like, 'We save everybody's life. Just not our own.'"
	Shared experience	"They need someone to relate. Again, they don't need to be an ex-drug addict; they don't need to be straight from the penitentiary; they don't need to be ANY of those things. Just be able to relate."
Wake-up call	Motivation for change	"If this didn't happen and, the way that I dealt with myself after it happened, if I didn't take that time to really put in work and to get myself to the point I am now, then I probably would have been in the streets with a gun right now, doing something ridiculous."
Positive person	Social support	"To have a positive person in your life can make a world of difference in it supports your belief in yourself and what you stand for, what you're tryin' to achieve. It's that affirmation that you can go on, despite tragedy and trauma. Now it's like your spirit is bein' ministered to. And, with anything that you do in life, you have to surround yourself with people who believe in you."

Institutional Mistrust

The context of "street culture" infused the injury and healthcare narratives of all participants to varying degrees. In particular, suspicion of dominant-culture institutions, such as police and healthcare, was common. The mistrust was exemplified by interactions with the police while receiving medical

care for the injury. The participants saw that healthcare personnel (e.g., ambulance, emergency department, or inpatient setting) would allow the police to question them while getting treatment. The police interrogation during medical treatment made some feel as if they were being treated as the perpetrator, and increased their sense of vulnerability. According to one participant, "…the only visitors I had was the police, comin' and talk to me. . . I felt like somebody was tryin' to murder ME, and I'm in jail, havin' all these cops comin' to see me and talk to me like (I'm the suspect in) a murder investigation."

Because this line between police and healthcare institution was blurred, in at least one case, it led to minimization of symptoms so that the healthcare personnel would not report this to the police. This participant discussed his presentation for treatment for a gun shot, "I just cut it short 'cause I wasn't too comfortable at tellin' 'em how I felt because if I tell 'em too much… next thing you know, the police is comin,' knockin' at my door."

Some of the institutional mistrust was based on a perception that healthcare clinicians are more motivated by money than by performing appropriate care, "… all the doctors and nurses and people in the healthcare field, when they first started out, they had genuine intentions. But, as years go on, big business corrupt their mind."

One manifestation of this mistrust was an unwillingness to give out accurate contact information to healthcare personnel. One participant explained this as, "the G-code," meaning gangster-code, a behavior in which the patient will be polite but not actually engage in the care, e.g. not give phone number, not attend follow-up appointments.

Foreshortened Future

Some participants reported that they had expected they would die at an early age. These feelings were reinforced by peers who were injured or died, and messages from adults, "A lot of my friends, they died when they were 15, 18 and, our teachers used to (say), "You guys, ya'll aint never gonna make it past 18 years old."

The injury reinforced expectations that getting shot or stabbed was inevitable. One participant said, "I left the hospital feeling like what happened to me is not a tragedy. It's just a way of life.'"

Participants heavily involved in street culture reported embracing this fate, "We had a gung-ho attitude about it. It's guns, violence, drugs. . . Today might be my day, so live it to the fullest." This apparent attitude suggests that clinicians appealing to self-preservation instincts in patients as motivation for treatment might not be successful.

Self-Reliance

Participants expected that individuals and families rely on themselves for help. "As a young Black male… I'm not one to go to somebody and say,…'I need some help, I need to talk." I rather. . . try to find out what to do, on my own. Get rid of my own stress." In particular, discussions of treatment for mental health issues were peppered with the language of stigma; one compared mental health treatment to putting a pacifier in a baby's mouth.

Others felt that mental health treatment is ineffective, partly because the therapist does not know the patient. One participant noted, "…spillin' your beans about everything personal to somebody who barely even knows you, and is giving you feedback and telling you what to do about YOUR situations…just 'cause it's their job."

Participants discussed forms of self-treatment, such as heavy use of substances, particularly alcohol and marijuana. "At times I cried, and then I started. . . breakin' up with my family, started usin' more drugs from alcohol to messin' with cocaine, with weed, then from there to messin' with heroin. . . to self-medicate myself. I figured if I self-medicate myself, I won't be in that mind."

Although 15 participants reported hazardous alcohol use or regular marijuana use, many reported that if they had a problem, they could stop on their own "My addiction is not so uncontrollable. . . I think my addiction is not worse than. . . quote/unquote 'hard core addicts'. . . I'm pretty much in control most of the time, when I make these bad decisions." While some participants did report mental health symptoms and a perceived need for treatment, few reported actively seeking or having received therapy. Some reported that they would have liked (or even now would like) therapy, but did not know how to obtain such help.

Logistical Issues

Several logistical barriers to obtaining medical and mental healthcare were noted. Participants discussed a lack of safety they felt after the injury, "I continued to work, but then I was always in the house, so it was like, I was hidin,' like I was a criminal…. And that's why, eventually, I just saved my money and moved on outside of the city."

Fear of public venues, such as public transportation, was common. This was particularly intense for those with newly acquired disabilities, including post-traumatic stress symptoms. "They wanted me to come to court and testify against him, … but, …after somethin' like this happen, there's no way you expect me just to get on the regular public transportation and come on down to the courthouse… That was like a death trap right there." Money struggles were common in the participants, including difficulty obtaining safe and affordable housing, and lack of employment.

Facilitators to Care

Desired Clinician Characteristics

All participants were specifically asked about racial and gender preferences for treating physicians and mental health professionals. None of the participants reported a need for the treating physician to be black. "If you're white, black, green, yellow, orange, as long as you're professional and you know how to do your job. . . It doesn't matter about race. I didn't have no black people on my team. A straight cracker saved my life." Medical competence, professionalism, and an "open-minded personality" that conveyed respect were cited as desirable characteristics of a clinician.

However, three did report wanting a mental health clinician to be black. One participant declared, "…she or he would definitely have to be the same color, hopefully, or

nationality… I'd definitely feel more comfortable with somebody in my, [background]." Five said they preferred clinicians to be one gender over another, with mixed opinions about which gender. Despite only a few requesting a black clinician, another eight participants wanted a mental health clinician who could relate to their experiences in the street, "It'd be nice to have someone who has experience,. . . like, they can say, 'Oh, I've been there,' or, 'I know someone who has been there.'"

Turning Point

Participants did report experiences and factors which promoted positive engagement with healthcare clinicians and motivated positive life changes. Despite feelings of a foreshortened future described by many participants, a number spoke about experiences which caused them to re-examine their lives. For some, the injury itself was a turning point in their life. It gave them time to think about past and future, and to see how their own actions might place them at risk. As one explains, "I was thinkin' about my future. I'm thinkin' about not doing the things that I had done in the past–sellin' drugs, doing drugs, goin' to jail, doing stupid stuff in the streets. . . God was tellin' me this was somethin' that may happen. It was a wake up call for me." More commonly the birth of a child or recognition of their importance to their children inspired personal change, "when (the baby) is yours and you're sittin' there, you're like, "Um, naw… I can't do this [street life] no more. I can't live like that."

Positive People

A number of participants used the phrase "positive person" to describe people who helped them with their mental recovery from the trauma (or other issues). These positive people included friends, neighbors, family members, bosses, religious communities, and a barber. Positive people focused on moving in a future direction, rather than dwelling in the past. It included people who spoke honestly and respectfully with the participant about their problems. According to one, "They need to be around a positive person, like a friend, and let them know like, 'Look, you need help, you need to, see somebody.'"

DISCUSSION

Among a group of black male victims of violence, the social environment strongly influenced engagement with medical and mental healthcare after an injury. These men perceived the importance of the medical treatment for their trauma. However, it did neither appear to integrate their reactions to the injury nor did it fully address the daily challenges they faced including mental health symptoms, disability, and safety concerns after the injury. Additionally, an underlying suspicion of institutions may interfere with taking advantage of potential resources of care. Beliefs about treatment for mental health and substance abuse may also have prevented them from accessing professional help. Engaging participants with supportive and forward social networks may best facilitate recovery.

The first sign of disengagement to healthcare may start at the moment of injury, with peritraumatic dissociation, a mental disconnect from the injury documented in studies of other violent injuries.[18,19] Just after the injury, thoughts centered on anger at the perpetrator and the circumstances of the injury. These two phenomena, and the foreign environment of the healthcare setting, may contribute to the patient being physically present but mentally and emotionally disconnected from the medical and surgical care. The narratives suggest that this chasm between the patient experience and medical care delivery has the potential to result in poorly planned discharges and preventable readmissions. While some factors may be beyond the control of medical personnel, knowledge of potential patient concerns, and challenges can facilitate development of potential responses.

While mistrust of healthcare institutions among black populations has been documented previously,[20–22] it has unique implications for victims of stabbing and shooting. Within inner-city black communities, suspicion of mainstream institutions may be explained in part by what sociologist Elijah Anderson calls, "The Code of the Street."[23] According to Anderson, this Code is a set of rules in response to the perceived failure of mainstream institutions to serve the needs of inner-city communities. The Code encourages an attitude of self-reliance, such as self-defense through neighborhood or gang affiliations, and aversion to asking for help, an attitude that may follow these youth into a healthcare setting. Cooperating with police may be interpreted as defying the Code and be a risk for further injury. Venkatesh,[24] in his book *Gang Leader for a Day*, substantiates the theme that poor black urban populations experience mainstream institutions as not serving their needs. Patients may view medical institutions as an extension of the legal system. For example, routine interviewing of patients by police detectives occurs during receipt of emergent treatment in both ambulances as well as Emergency Departments (EDs). Given this experience, patients may not view the postinjury care experience as a secure and trusting environment, rather- it has potential for worsening safety after discharge.

Emergency medical services should consider instituting policies about asking patient permission before allowing police to question patients. While this may not be possible before entering hospital EDs, instituting such policies could be considered in hospital EDs, in the same way that patients may refuse other visitors. In addition, medical personnel should clearly delineate what patient information is available to police (or others) and what is confidential.

Physical disabilities stemming from the injury increase the vulnerability already engendered by injury itself and subsequent encounters with police. This may lead to a decreased sense of safety, particularly heightened in venues such as public transportation. The fear of public transportation may be a major obstacle for these men to obtain postinjury care and to take advantage of available resources. Providing patients with private and safe transportation may help to improve follow-up care.

Appropriate and safe housing is also a challenge for a number of these men. Prior housing arrangements may not be appropriate due to physical limitations (e.g. not wheelchair accessible). They also may not feel safe from either gang-

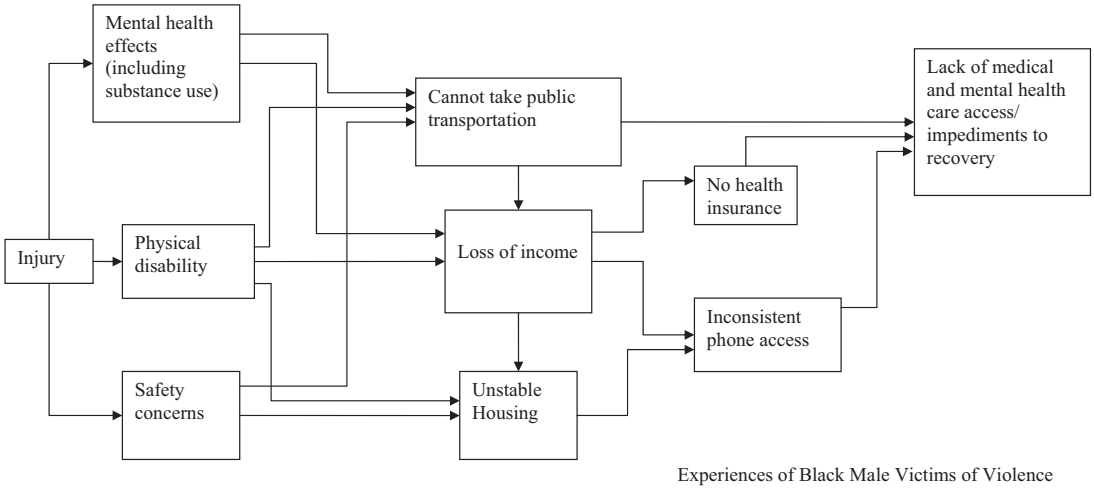


Figure 1. Postinjury barriers to medical and mental healthcare for black male victims of community violence.

related retaliation (to keep victim from disclosing to police), high family tensions, difficulty paying rent due to loss of income, or posttraumatic stress symptoms from fear of injury in the same neighborhood. These survival issues compete with medical needs, especially in a group of young men who hold self-reliance as a core value. Future programs on preventing further postinjury morbidity should consider providing supported and safe housing to allow the participant to recover. Figure 1 depicts a theoretical model of how logistical barriers stemming from injury may affect access to healthcare.

The easy availability of substances, particularly alcohol and marijuana, provided an escape from the emotional toll of stabbing and shooting for some study participants. As noted by Rich and Grey,[25] substance use is one link on the pathway to re-injury. Getting professional help for substance use or mental health symptoms would entail overcoming a variety of barriers for black male violence victims. The nature of treatment involves asking for help, acknowledging weakness and trusting of mainstream institutions, which contradicts a core value of self-reliance. Inaccurate information about effectiveness of treatment contributes to these barriers. Contradictorily, most study participants desired relief from mental suffering but did not know how to access such help, suggesting that a public health campaign about substance abuse and mental health treatment, along with expanded access, may decrease some barriers.

The power of the social network suggests another route that medical clinicians might draw on: working (with patient permission) on engaging a circle of positive people in the recovery plan. This may help encourage patients to engage in self-care and use the available treatments.

Limitations

Qualitative studies, by design, are meant to generate ideas and not provide conclusions generalized to other populations. A common drawback of qualitative research is related to bias, including bias of the interviewers who may have unknowingly influenced participant responses and bias of the researchers analyzing the data. To counter those biases, interviewers were extensively trained. As well, a large diverse, interdisciplinary team performed the analyses and formulated the conclusions with participation from all team members. It is not known whether this sample differed greatly from those unwilling to participate in research projects. Additionally, the participants' memories may have evolved since the actual events took place, and thus, the reporting may not reflect actual experiences. While interviews were scheduled with black male interviewers, participants who arrived at unscheduled times and did not have adequate phone access to reschedule were interviewed by available trained personnel. It is not clear how this affected study results.

CONCLUSIONS

Effective emergency care for black male victims of stabbing and shooting should go beyond stopping the acute bleeding, and encompass methods to build bridges to patients' needs. This would include delineating clear boundaries between patient rights and law enforcement needs, and then assessing for safety, transportation, and housing issues after discharge. Campaigns to educate and decrease stigma of mental and substance abuse care combined with facilitated access to such treatment may also improve outcomes.

REFERENCES

1. National Center for Injury Prevention and Control. *Leading Causes of Death Reports*. Atlanta, GA: Center for Disease Control and Prevention; 2005.
2. Fox J, Swatt M. The Recent Surge in Homicides involving Young Black Males and Guns: Time to Reinvest in Prevention and Crime Control. Boston, MA: Northeastern University; 2008:1–8.
3. Holbrook TL, Hoyt DB, Coimbra R, Potenza B, Sise M, Anderson JP. Long-term posttraumatic stress disorder persists after major trauma in adolescents: new data on risk factors and functional outcome. *J Trauma.* 2005;58:764–769; discussion 769–771.

4. Adesanya AA, da Rocha-Afodu JT, Ekanem EE, Afolabi IR. Factors affecting mortality and morbidity in patients with abdominal gunshot wounds. *Injury.* 2000;31:397–404.
5. Becker MG, Hall JS, Ursic CM, Jain S, Calhoun D. Caught in the Crossfire: the effects of a peer-based intervention program for violently injured youth. *J Adolesc Health,* 2004;34:177–183.
6. Cooper C, Eslinger DM, Stolley PD. Hospital-based violence intervention programs work. *J Trauma-Injury Infect Crit Care.* 2006;61:534–537; discussion 537–540.
7. Zun LS, Downey L, Rosen J. The effectiveness of an ED-based violence prevention program. *Am J Emerg Med.* 2006;24:8–13.
8. Shibru D, Zahnd E, Becker M, Bekaert N, Calhoun D, Victorino GP. Benefits of a hospital-based peer intervention program for violently injured youth. *J Am Coll Surg.* 2007;205:684–689.
9. Zun LS, Downey LV, Rosen J. Violence prevention in the ED: linkage of the ED to a social service agency. *Am J Emerg Med.* 2003;21:454–457.
10. Smedley B, Stith A, Nelson A. *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care*. Washington, DC: Institute of Medicine; 2002.
11. Ravenell JE, Whitaker EE, Johnson WE Jr. According to him: barriers to healthcare among African-American men. *J Natl Med Assoc.* 2008; 100:1153–1160.
12. Strauss A, Corbin J. *Basics of Qualitative Research: Grounded Theory Procedures and Techniques.* Newbury Park, CA: SAGE Publications; 1990:270.
13. Weathers FW, Litz BT, Herman JA, Huska JA, Keane TM. The PTSD Checklist (PCL): reliability, validity and diagnostic utility. In: Annual Conference of the ISTSS, San Antonio, TX, 1993.
14. Weathers FW, Litz BT, Huska JA, Keane TM. *PCL-C for DSM*, IV ed. Boston, MA: N.C.F.P.B.S. Division; 1994.
15. Blanchard EB, Jones-Alexander J, Buckley TC, Forneris CA. Psychometric properties of the PTSD checklist (PCL). *Behav Res Ther.* 1996; 34:669–673.
16. 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;88:791–804.
17. *NVIVO version7*. Victoria, Australia: QSR International; 2006.
18. Fein JA, Kassam-Adams N, Vu T, Datner EM. Emergency department evaluation of acute stress disorder symptoms in violently injured youths. *Ann Emerg Med.* 2001;38:391–396.
19. Marshall GN, Schell TL. Reappraising the link between peritraumatic dissociation and PTSD symptom severity: evidence from a longitudinal study of community violence survivors. *J Abnorm Psychol.* 2002;111: 626–636.
20. Armstrong K, McMurphy S, Dean LT, et al. Differences in the patterns of health care system distrust between blacks and whites. *J Gen Intern Med.* 2008;23:827–833.
21. Boulware LE, Cooper LA, Ratner LE, LaVeist TA, Powe NR. Race and trust in the health care system. *Public Health Rep.* 2003;118:358–365.
22. Halbert CH, Armstrong K, Gandy OH Jr, Shaker L. Racial differences in trust in health care providers. *Arch Intern Med.* 2006;166:896–901.
23. Anderson E. The code of the streets. *Atlantic Monthly.* 1994;273:81–94.
24. Venkatesh S. *Gang Leader for a Day.* London, England: Penguin Press; 2008.
25. Rich JA, Grey CM. Pathways to recurrent trauma among young black men: traumatic stress, substance use, and the "code of the street." *Am J Public Health.* 2005;95:816–824.

LESS IS MORE

Proton Pump Inhibitors and Risk for Recurrent *Clostridium difficile* Infection

Amy Linsky, MD; Kalpana Gupta, MD, MPH; Elizabeth V. Lawler, DSc; Jennifer R. Fonda, MA; John A. Hermos, MD

Background: Proton pump inhibitors (PPIs) are widely used gastric acid suppressants, but they are often prescribed without clear indications and may increase risk of *Clostridium difficile* infection (CDI). We sought to determine the association between PPI use and the risk of recurrent CDI.

Methods: Retrospective, cohort study using administrative databases of the New England Veterans Healthcare System from October 1, 2003, through September 30, 2008. We identified 1166 inpatients and outpatients with metronidazole- or vancomycin hydrochloride–treated incident CDI, of whom 527 (45.2%) received oral PPIs within 14 days of diagnosis and 639 (54.8%) did not. We determined the hazard ratio (HR) for recurrent CDI, defined by a positive toxin finding in the 15 to 90 days after incident CDI.

Results: Recurrent CDI was more common in those exposed to PPIs than in those not exposed (25.2% vs 18.5%). Using Cox proportional survival methods, we determined that the adjusted HR of recurrent CDI was greater in those exposed to PPIs during treatment (1.42; 95% confidence interval [CI], 1.11-1.82). Risks among exposed patients were highest among those older than 80 years (HR, 1.86; 95% CI, 1.15-3.01) and those receiving antibiotics not targeted to *C difficile* during follow-up (HR, 1.71; 95% CI, 1.11-1.64).

Conclusions: Proton pump inhibitor use during incident CDI treatment was associated with a 42% increased risk of recurrence. Our findings warrant further studies to examine this association and careful consideration of the indications for prescribing PPIs during treatment of CDI.

Arch Intern Med. 2010;170(9):772-778

Author Affiliations: Section of General Internal Medicine, Department of Medicine, Boston Medical Center (Drs Linsky and Hermos), Massachusetts Veterans Epidemiology Research and Information Center, Veterans Affairs Cooperative Studies Program, Veterans Affairs Boston Healthcare System (Drs Gupta, Lawler, and Hermos and Ms Fonda), Section of Infectious Diseases (Dr Gupta), Department of Medicine (Dr Hermos), Boston University School of Medicine, and Department of Medicine, Harvard Medical School (Dr Lawler), Boston, Massachusetts.

PROTON PUMP INHIBITORS (PPIs) are among the most commonly prescribed agents in outpatient and inpatient settings, but they are often used without a clear indication.[1,2] Although PPIs have been considered relatively safe, studies have demonstrated that PPI use and the use of other gastric acid suppressants are associated with a variety of infections, including *Salmonella* and *Campylobacter* gastroenteritis and community-acquired and nosocomial pneumonia.[3-7]

See also pages 747, 749, 751, 765, 779, and 784

Proton pump inhibitors have also been associated with *Clostridium difficile* infection (CDI).[8-11] *Clostridium difficile* is a spore-forming bacterium with a toxin that is associated with diarrheal illnesses. Symptomatic recurrence occurs in approximately 20% of patients after treatment, with even higher rates after subsequent episodes.[12-15] The incidence of CDI in the United States has nearly tripled in the past decade, with antibiotic and hospital exposures recognized as the preeminent risks for infection.[16,17] Other putative risk factors include older age, frailty, and environmental factors.[18]

Several observational studies have associated community-acquired and nosocomial CDI with PPI use, finding risks of CDI 2 to 3 times higher in PPI users compared with nonusers.[8-11,17,19,20] Other research has failed to associate PPIs with CDI.[21,22] The exact relationship between PPI use and incident CDI remains elusive, and no causative pathway has been demonstrated. The relationship between PPI exposure and the risk of recurrent CDI is even less clear, with fewer studies and conflicting findings.[23,24] To further address this question, we conducted a retrospective cohort analysis using data from the New England Veterans Healthcare System (VISN 1) to determine the association between PPI use and the risk of recurrent CDI.

METHODS

STUDY SETTING

We used linked pharmacy and administrative databases from VISN 1 from October 1, 2003, through September 30, 2008. The New England region consists of 8 Veterans Affairs (VA) medical centers and affiliated clinics, including 3 large, urban, tertiary care centers and 5 smaller facilities providing inpatient and outpatient care. The VISN 1 pharmacy files were obtained from Veterans Health Information Systems and Technology Architecture. The data elements of the outpatient and inpatient pharmacy files used in this study include patient identification, date of birth, drug name and dose, drug administration route, quantity, VA drug class, date of original prescription, number of days of medication supplied, and refill date. Laboratory data were obtained from the electronic medical record. Comorbidities were captured via the VA national patient care database located at the Austin Automation Center, Austin, Texas. Analyses were conducted at the Massachusetts Veterans Epidemiology Research and Information Center, VA Cooperative Studies coordinating center, VA Boston Healthcare System. The study was approved by the institutional review board of the VA Boston Healthcare System, which waived informed consent and Health Insurance Portability and Accountability Act of 1996 authorization.

STUDY POPULATION

From the data sets, we identified patients with a first positive finding of a *C difficile* toxin from October 1, 2004, through September 30, 2008. This search yielded the index date for 1549 patients. The laboratories of VISN 1 use enzyme immunoassay tests for toxins A and B, with a standard practice of testing only loose stool samples. This sample of incident cases was then restricted to 1546 patients with a minimum of 1 year of prior VA health care system use to establish their use of the VA system. We further limited the study population to 1408 patients for whom treatment was initiated within the VA system with metronidazole or oral vancomycin hydrochloride and then to 1166 patients whose treatment started within 3 days before or after the index CDI (**Figure 1**).

PRIMARY EXPOSURE AND OUTCOMES

The main exposure variable was the use of any oral PPI during the 14 days after incident CDI diagnosis, coinciding with the treatment window. For inpatients (n=981), PPI exposure was defined as any pharmaceutical dispensing of a PPI during this postdiagnosis window. For outpatients (n=185), the end date of the most recent antecedent prescription was extended by 10% (eg, 9 days for a 90-day prescription) to account for potentially missed doses. If the predicted prescription end date fell after the index date, then the subject was placed in the PPI-exposed group. For all outpatients, actual prescription end dates fell beyond the index date. With these exposure criteria, 527 patients (45.2%) were categorized as PPI exposed and 639 patients (54.8%) as non–PPI exposed (Figure 1).

The primary outcome measure was a positive finding for a *C difficile* toxin occurring 15 to 90 days after the incident CDI diagnosis date. We established a 15- to 90-day follow-up window to provide a period of observation that would capture the greatest number of possible recurrent cases based on clinical observations and previous studies.[12,24] Patients were censored at death or 90 days after the index date. We also calculated the time to recurrence, defined as the number of days from the later of either the index date or the start of incident CDI treatment until the recurrent diagnosis date. Thus, the calculated time to recurrence could be as short as 12 days.

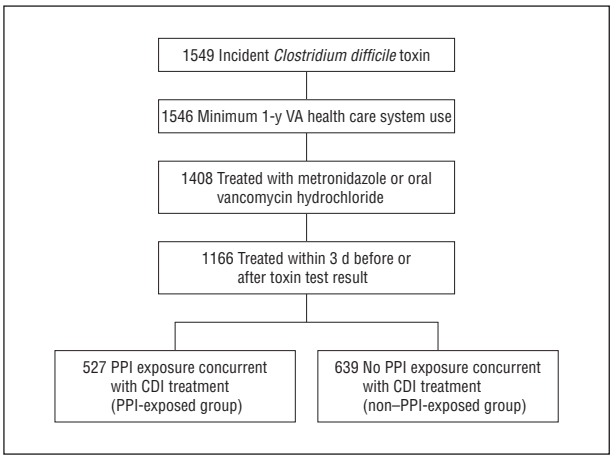


Figure 1. Determination of study population. CDI indicates *Clostridium difficile* infection; PPI, proton pump inhibitor; and VA, Veterans Affairs.

COVARIATES

Covariates that may influence the risk of recurrent CDI and those that may influence exposure to PPIs were included in our analysis. These variables were age, sex, comorbid conditions, medication used before the index date, initial incident CDI antibiotic treatment (metronidazole or vancomycin), non–CDI-targeted antibiotic exposure during follow-up, VA nursing home admission during the study period, and, for those who were inpatients, the duration of hospitalization after the index date.

Comorbidities were determined from administrative records using *International Classification of Diseases, Ninth Revision, Clinical Modification* (*ICD-9-CM*) codes, including selected diagnoses recorded in the 2 years before the index date. Comorbidities and their *ICD-9-CM* codes were hypertension (codes 401-404), diabetes (250), ischemic heart disease (410-414), chronic obstructive pulmonary disease (491, 492, and 496), esophageal disease (530), peptic ulcer disease (531-535), solid tumors (140-199), and rheumatologic disease (710, 714, and 720). Medication used in the 90 days before the index date was included to represent the acuity and severity of the comorbid conditions and those drugs that might directly increase susceptibility to infections. Specific medications were antibiotics, systemic and inhaled corticosteroids, chemotherapeutics, and immunomodulators. In addition, we recorded inpatient and outpatient prescriptions for H_2 receptor antagonists.

We determined the initial antibiotic treatment for CDI because differences in clinical severity may affect the choice of antibiotic. There may also be different treatment success rates.[25,26] We evaluated how many patients changed from metronidazole to vancomycin therapy within the 14 days after diagnosis. Because antibiotic exposure is a major risk factor for CDI,[20,27,28] we determined whether patients were exposed to non–CDI-targeted antibiotics during follow-up. Similarly, because hospital exposure is a major risk factor,[29,30] we determined the number of days patients were hospitalized after incident CDI diagnosis. We categorized this variable as 0 days, discharged fewer than 14 days after the index diagnosis, and discharged at least 14 days after the index diagnosis. To account for the possible effect of PPI exposure from day 15 to censorship, we determined whether patients were dispensed or prescribed PPIs by using the same criteria as during the treatment window.

Table 1. Characteristics of the Study Population[a]

Characteristic	PPI Exposure Concurrent With CDI Treatment (n=527)	No PPI Exposure Concurrent With CDI Treatment (n=639)	*P* Value
Patient demographics			
Age, median (IQR), y	74 (63-81)	74 (63-82)	.31
Men	507 (96.2)	626 (98.0)	.07
Patient comorbidities			
Hypertension	463 (87.9)	545 (85.3)	.20
Diabetes mellitus	247 (46.9)	276 (43.2)	.21
Ischemic heart disease	329 (62.4)	319 (49.9)	<.001
Chronic obstructive pulmonary disease	275 (52.2)	289 (45.2)	.02
Esophageal disease	273 (51.8)	232 (36.3)	<.001
Peptic ulcer disease	144 (27.3)	137 (21.4)	.02
Solid tumor	220 (41.7)	243 (38.0)	.20
Rheumatologic disease	55 (10.4)	38 (5.9)	.005
Chronic kidney disease	95 (18.0)	96 (15.0)	.17
Medication use in 90 d before index date			
Antibiotics	470 (89.2)	567 (88.7)	.81
H_2 receptor antagonists	133 (25.2)	170 (26.6)	.60
Systemic corticosteroids	155 (29.4)	146 (22.8)	.01
Inhaled corticosteroids	60 (11.4)	73 (11.4)	.98
Chemotherapy	11 (2.1)	14 (2.2)	.90
Immunosuppressants	10 (1.9)	6 (0.9)	.16
Incident case initial antibiotic treatment			
Metronidazole	504 (95.6)	600 (93.9)	.19
Oral vancomycin hydrochloride	23 (4.4)	39 (6.1)	
Time to discharge from index date, d			
0 (outpatients)	68 (12.9)	117 (18.3)	.007
<14	196 (37.2)	330 (51.6)	
≥14	263 (49.9)	192 (30.0)	
VA nursing home admission during days 1-90	47 (8.9)	67 (10.5)	.37
Non-CDI antibiotic exposure during days 15-90	220 (41.7)	206 (32.2)	<.001
PPI exposure during days 15-90	434 (82.4)	41 (6.4)	<.001

Abbreviations: CDI, *Clostridium difficile* infection; IQR, interquartile range; PPI, proton pump inhibitor; VA, Veterans Affairs.

[a]Unless otherwise indicated, data are expressed as number (percentage) of patients.

STATISTICAL ANALYSIS

We used χ^2 tests to compare categorical variables between PPI-exposed and non–PPI-exposed groups, setting $\alpha = .05$. We compared age using Wilcoxon 2-sample tests. Rate ratios and 95% confidence intervals (CIs) were determined for recurrence during the follow-up period. To account for time to recurrence, we used survival analysis techniques, calculating Cox proportional hazards ratios (HRs) for additional positive findings of *C difficile* toxin. We included in the final model those variables that were statistically different between the groups or that were thought to be clinically relevant. We tested the proportional hazards assumption by including an interaction term in the final model between PPI exposure and the natural logarithm of time. Post hoc stratified analyses were conducted by patient age, non-CDI antibiotic exposure, and PPI exposure after treatment for incident CDI.

SENSITIVITY ANALYSIS

Positive toxin findings alone may not represent clinically relevant, recurrent CDI. Thus, we determined HRs excluding 34 patients (13.5% of 251 total cases of recurrence) with a new positive toxin finding for whom there was no VA pharmacy record of additional CDI antibiotic treatment. Because continued hospital exposure could favor CDI recurrence, we determined whether alternative categorization of inpatient time to discharge influenced hazard rates. We also analyzed to what extent patients with recurrent CDI had intervening positive toxin results. All analyses were performed using SAS statistical software (version 9.1.3; SAS Institute Inc, Cary, North Carolina).

RESULTS

STUDY POPULATION

The study population of 1166 veteran patients meeting criteria for recurrent CDI was predominantly male (97.2%), with a median age of 74 (interquartile range, 63-82) years. Overall, 527 (45.2%) of the patients were exposed to a PPI concurrent with antibiotic treatment (PPI-exposed group) and 639 (54.8%) were not exposed to concurrent PPI use (non–PPI-exposed group) (Figure 1 and **Table 1**). Almost all PPI-exposed patients (96.7%) were prescribed omeprazole, 20 mg once daily, or the equivalent dose of another PPI; the remaining patients received lower or higher PPI doses. Age and sex were similar among the PPI-exposed and non–PPI-exposed patients. The PPI-exposed group had a higher prevalence of *ICD-9-CM* coded ischemic heart disease, chronic obstructive pulmonary disease, esophageal disease, peptic ulcer disease, and rheumatologic disease. This group also had greater exposure to systemic corticosteroids. Metronidazole was the initial antibiotic used for treatment of incident CDI in more than 90% of both groups. Similar proportions (11%) in each group were switched from metronidazole to oral vancomycin within the treatment window. However, patients in the PPI-exposed group were more likely to have been inpatients and were more likely to have inpatient stays exceeding 14 days than those in the non–PPI-exposed group. Also, those in the PPI-exposed group were more likely to have been exposed to non-CDI antibiotics during follow-up. Of those in the PPI-exposed group, 434 (82.4%) had continued exposure to PPIs during follow-up days 15 to 90. In the non–PPI-exposed group, there were only 41 patients (6.4%) exposed to PPIs during this follow-up. Similar proportions of each group had an admission to a VA nursing home during the study window.

RELATIONSHIP OF PPI TO RECURRENT CDI

The primary outcome of additional positive findings for *C difficile* toxin occurred in 251 patients (21.5%). The unadjusted incidence of recurrent toxin in the 15 to 90 days after incident CDI was greater in the PPI-exposed compared with the non–PPI-exposed group (133 of 527 [25.2%] vs 118 of 639 [18.5%]). Unadjusted Kaplan-Meier recurrence-free survival curves are seen in **Figure 2**, with an associated HR of 1.42 (95% CI, 1.11-

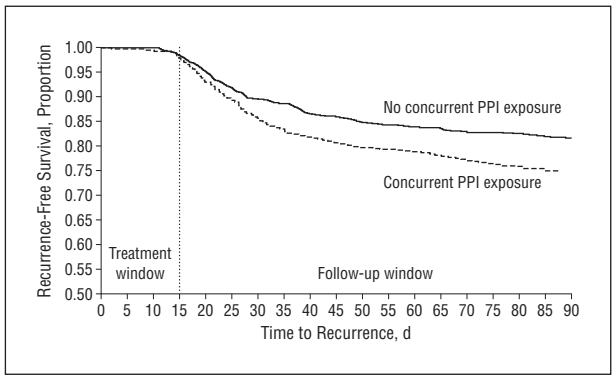


Figure 2. Recurrence-free survival in those exposed vs unexposed to proton pump inhibitors (PPIs) during treatment for incident *Clostridium difficile* infection. Time to recurrence started from the incident toxin finding or the start of antibiotic treatment (≤3 days after the diagnosis).

1.82; *P*=.006) for those in the PPI-exposed group compared with those in the non–PPI-exposed group. After adjusting for age, initial incident CDI antibiotic treatment (metronidazole or vancomycin), additional antibiotic exposure (yes or no), duration of hospital exposure (0, <14, or ≥14 days), and differing baseline comorbidities (ischemic heart, esophageal, peptic ulcer, pulmonary, and rheumatologic diseases) and medications (systemic corticosteroids), the HR associated with PPI exposure remained elevated at 1.42 (95% CI, 1.10-1.83; *P*=.008) (**Table 2**). The proportional hazards assumption was met for the adjusted model because the interaction term between exposure and the logarithm of time was not significant (*P*=.65).

EFFECT MODIFICATION

We evaluated potential effect modification by age, non-CDI antibiotic exposure during follow-up, and PPI exposure during follow-up. We found an increasing risk of recurrence associated with PPI use with increasing age (Table 2). Stratification by non-CDI antibiotic exposure in the follow-up window modified the association of PPIs and recurrent CDI on the ratio scale. Those exposed to additional antibiotics had a 71% greater risk of recurrence associated with PPIs, whereas those not exposed to antibiotics had a 30% greater risk of recurrence. However, we were not sufficiently powered to detect a statistically significant effect modification for age or non-CDI antibiotic exposure.

Most patients in the PPI-exposed group (434 of 527 [82.4%]) and very few of the non–PPI-exposed group (41 of 639 [6.4%]) received PPIs during the 15- to 90-day follow-up (Table 1). Patients prescribed PPIs during both treatment and follow-up had an increased risk of recurrence compared with those who were not exposed to PPIs in either time window (HR, 1.44; 95% CI, 1.09-1.89; *P*=.01). Greater risk for recurrence, although not powered for statistical significance, was also seen in the smaller subgroups who were exposed to PPIs in the treatment period only (41 patients; HR, 1.30; 95% CI, 0.68-2.51; *P*=.43) or in the follow-up period only (93 patients; HR, 1.51; 95% CI, 0.93-2.46; *P*=.10).

Table 2. Association of CDI Treatment-Concurrent PPI Exposure With Recurrent CDI Within 90 Days

Model	HR (95% CI)	*P* Value
Unadjusted	1.42 (1.11-1.82)	.006
Adjusted[a]	1.42 (1.10-1.83)	.008
Age stratified, y[a]		
<60 (n=189)	1.19 (0.56-2.55)	.65
60-80 (n=593)	1.32 (0.94-1.85)	.11
>80 (n=384)	1.86 (1.15-3.01)	.01
Non-CDI antibiotic exposure stratified[a]		
Antibiotic exposure (n=426)	1.71 (1.11-2.64)	.01
No additional antibiotic exposure (n=740)	1.30 (0.94-1.79)	.12

Abbreviations: CDI, *Clostridium difficile* infection; CI, confidence interval; HR, hazard ratio; PPI, proton pump inhibitor.

[a] Adjusted for age, incident CDI treatment, additional antibiotic exposure, length of hospital exposure, ischemic heart disease, esophageal disease, rheumatologic disease, peptic ulcer disease, pulmonary disease, and systemic corticosteroid use.

SENSITIVITY ANALYSES

We determined the proportion of those classified as having recurrent CDI whose treatment with CDI-targeted antibiotics was documented in the VA pharmacy. Of the 251 cases meeting our criteria for recurrent CDI, 217 (86.5%) were prescribed metronidazole or oral vancomycin at the time of the additional positive toxin finding. Restricting our risk analyses to these patients did not change our estimates (adjusted HR, 1.52; 95% CI, 1.15-2.01; *P*=.003).

Because a change from metronidazole to vancomycin therapy may reflect more severe disease clinically or failure to respond to the original CDI treatment antibiotic, we evaluated the influence of any exposure to vancomycin on the association of PPIs and recurrence. Controlling for any vancomycin treatment within the first 14 days did not alter our findings (adjusted HR, 1.42; 95% CI, 1.10-1.84; *P*=.008).

We conducted adjusted analyses using varied categorical lengths of hospitalization following incident CDI diagnosis and found no differences in the HRs associating PPI exposure with recurrent CDI (data not shown).

Of the 251 recurrent cases, 121 (48.2%) had no intervening test for *C difficile* toxin, 100 (39.8%) had an intervening test with a negative result, and 30 (12.0%) had an intervening test with positive results, mostly within several days of the index date. There was no difference in PPI exposure among those with intervening negative test results or no intervening tests (*P*=.43).

COMMENT

Our findings indicate that PPI use concurrent with treatment for CDI was associated with a 42% increased risk of recurrent CDI in the subsequent 15 to 90 days. In stratified analyses, exposure during treatment and within the follow-up window was associated with a 44% greater recurrence, whereas patients prescribed PPIs in only 1 of the 2 exposure windows (ie, treatment only or follow-up only) demonstrated a higher but nonstatistically significant risk of recurrent infection.

Proton pump inhibitors have been linked to higher risks of community-acquired and nosocomial CDI,[8-11,19] but other studies have not shown this association.[21,22] Few studies have specifically focused on the association of PPIs with recurrent CDI. A retrospective medical record review of 140 patients in a single VA medical center demonstrated 4-fold increased odds of recurrence in those exposed to PPIs concurrently with CDI treatment, although HRs accounting for time to recurrence were not analyzed.[24] A small prospective medical record review of patients with endoscopically proved pseudomembranous colitis found higher relapse rates in those prescribed antiulcer medications but did not distinguish between PPIs and other gastric acid suppressants.[31] Similarly, a case-control study showed a slightly greater risk of recurrence in patients exposed to H_2 receptor antagonists, but the investigators did not assess the risk of PPIs.[13] Others have not found a difference in recurrence rates due to acid-suppressing medications[23]; however, those investigators did not distinguish between PPIs and H_2 receptor antagonists, which are generally less potent. The data presented herein represent, to our knowledge, the largest evaluation of this potentially modifiable exposure to PPIs and risk of CDI recurrence.

One potential mechanism to explain this association may be that elevated gastric pH levels facilitate the growth of potentially pathogenic upper and lower gastrointestinal tract flora.[3,32,33] Although *C difficile* spores are acid resistant, vegetative forms are susceptible to acidity.[34,35] Furthermore, elevated gastric pH levels may allow or facilitate conversion from spore to vegetative forms of *C difficile* in the upper gastrointestinal tract. Other potential mechanisms include impairment of leukocytes and other immune responses and antimicrobial properties of PPIs.[36-39]

Our study had several strengths, including use of data from the VA's large, multisite, integrated health care system. The ability to link information from laboratory, pharmacy, and administrative data enabled comprehensive assessment of multiple aspects of patients' health care. We also incorporated time to recurrence into our analyses, and stratified and sensitivity analyses supported the association of PPI use with recurrent CDI. Finally, our data are compatible with plausible biological mechanisms and with results from previous studies indicating increased risk of infections, including CDI, with PPI exposure.[4-11]

As with all studies using administrative and clinical data extracts, our findings should be viewed in the context of the following limitations. Use of observational databases allows for potential misclassification of exposure (eg, for outpatients, use of non-VA prescribed or over-the-counter PPIs or other gastric acid suppressants, and conversely, nonadherence to VA-prescribed PPI therapy). However, there exists significant financial incentive for most veterans to obtain prescriptions directly from the VA. Furthermore, because most of our sample initially had CDI treatment as inpatients, we are confident in the quality of our PPI exposure data.

A positive test result for a *C difficile* toxin by itself does not necessarily indicate a clinically relevant recurrence and could lead to potential misclassifications of the outcome. However, our findings remained robust when we assessed overdiagnosis of recurrent CDI by analyzing only the 217 subjects (86.5%) who had an additional positive toxin finding and additional documented treatment with CDI-targeted antibiotics. The hazards of recurrence in this restricted sample were similar to those of the entire study population (adjusted HRs, 1.52 vs 1.42, respectively).

Our data cannot, with certainty, determine to what extent the patients had recurrent CDI after a short period of treatment success or experienced treatment failure. Further examination revealed that 39.8% of the recurrent cases had an intervening negative stool toxin finding and 48.2% had no intervening tests. There was no difference in PPI exposure among these subgroups. From these data, there was little evidence to suggest that patients classified as having recurrent CDI had persistently positive toxin results between their index and recurrence dates. Nevertheless, to distinguish recurrent disease from treatment failure would require a prospective, systematic testing of subjects receiving treatment for CDI.

Patient frailty has been associated with CDI.[40] Although we controlled for differences in comorbid conditions, other than incorporating use of selected medications to treat these conditions, we did not directly assess the severity of illnesses. In addition, our study was conducted in a predominantly male veteran population, and our results may not generalize to all patients. Previous studies have found women to be at greater risk for CDI than men,[8] but other large investigations have associated PPIs with *C difficile* in both sexes.[10,40] Overall, these findings may be generalized to other elderly populations with similar access to PPIs.

After incident CDI diagnosis, our sample had other potentially important exposures increasing their risk of recurrent CDI. Although we controlled for patients' time in the hospital after incident disease, we were unable to incorporate *C difficile* pressure in specific settings such as hospital wards.[18] We also did not find any difference between groups for nursing home admissions during the study window. In our adjusted analyses, controlling for non-CDI antibiotic exposure during follow-up did not affect the risk associated with PPI exposure during treatment. When stratifying results, the risk associated with PPIs increased for those with additional antibiotic exposure, and statistical significance was no longer reached in those who received no further antibiotic treatment. Given the direction of the association, lack of further antibiotic treatment does not necessarily negate the effect of PPIs.

Most patients remained exposed or nonexposed to PPIs during the treatment and follow-up windows; very few received PPIs only during days 1 to 14 or only during days 15 to 90. Continued exposure to PPIs was associated with a risk of recurrent CDI similar to that found for the originally defined PPI-exposed group. Thus, to what extent the PPI exposure concurrent with CDI treatment, during follow-up, or during both periods is driving the association with recurrence is unclear.

The effect of any risk attributable to PPIs, regardless of magnitude, is increased owing to the high prevalence of use for this class of medication. Proton pump inhibi-

tors are often used without a clear indication. Studies of outpatients have found that of those receiving long-term acid-suppressive therapy, only 61% had a relevant gastrointestinal tract–related diagnosis.[2] Similarly, studies of inpatients found acid-suppressing medications frequently prescribed for stress ulcer prophylaxis in non–intensive care unit settings,[41-43] a practice contrary to the American Society of Health System Pharmacists guidelines.[44] Choudhry et al[45] studied patients with CDI and PPI use for prescription indication; an appropriate indication was not identified in 53.4% of the patients. We did not assess the indications for PPI use in our patients.

In conclusion, the study identified a 42% increased risk of recurrent CDI related to PPI use. Given the morbidity and cost associated with recurrent CDI and the lack of readily modifiable risk factors, our findings have important clinical implications. The data presented herein support the need for critical assessment of PPI use in patients being treated for CDI as well as further research to test this association.

Accepted for Publication: November 20, 2009.
Correspondence: Amy Linsky, MD, Department of General Internal Medicine, Boston Medical Center, 801 Massachusetts Ave, Second Floor, Boston, MA 02118 (Amy .Linsky@bmc.org).
Author Contributions: Dr Linsky had full access to all the data in the study and takes responsibility for the integrity and accuracy of the data analysis. *Study concept and design:* Linsky, Gupta, Lawler, and Hermos. *Acquisition of data:* Linsky and Fonda. *Analysis and interpretation of data:* Linsky, Gupta, Lawler, Fonda, and Hermos. *Drafting of the manuscript:* Linsky and Hermos. *Critical revision of the manuscript for important intellectual content*: Linsky, Gupta, Lawler, Fonda, and Hermos. *Statistical analysis:* Linsky, Lawler, and Fonda. *Administrative, technical, and material support:* Linsky, Lawler, Fonda, and Hermos. *Study supervision:* Linsky, Gupta, and Hermos.
Financial Disclosure: None reported.
Funding/Support: The study was supported by the resources of the VA Cooperative Studies Program and using the facilities of the VA Boston Healthcare System.
Previous Presentations: This study was presented at the 32nd Annual Meeting of the Society for General Internal Medicine; May 14, 2009; Miami, Florida; and at the 25th International Conference on Pharmacoepidemiology and Therapeutic Risk Management; August 16, 2009; Providence, Rhode Island.
Additional Contributions: Galina Sokolovskaya, MS, provided programming assistance and David Gagnon, MD, provided statistical consultation.

REFERENCES

1. Heidelbaugh JJ, Goldberg KL, Inadomi JM. Overutilization of proton pump inhibitors: a review of cost-effectiveness and risk. *Am J Gastroenterol*. 2009; 104(2)(suppl 2):S27-S32.
2. Jacobson BC, Ferris TG, Shea TL, Mahlis EM, Lee TH, Wang TC. Who is using chronic acid suppression therapy and why? *Am J Gastroenterol*. 2003;98(1): 51-58.
3. Brummer RJ, Stockbrugger RW. Effect of nizatidine 300 mg at night and omeprazole 20 mg in the morning on 24-hour intragastric pH and bacterial overgrowth in patients with acute duodenal ulcer. *Dig Dis Sci*. 1996;41(10):2048-2054.
4. Dial MS. Proton pump inhibitor use and enteric infections. *Am J Gastroenterol*. 2009;104(2)(suppl 2):S10-S16.
5. García Rodríguez LA, Ruigómez A, Panés J. Use of acid-suppressing drugs and the risk of bacterial gastroenteritis. *Clin Gastroenterol Hepatol*. 2007;5(12): 1418-1423.
6. Sarkar M, Hennessy S, Yang YX. Proton-pump inhibitor use and the risk for community-acquired pneumonia. *Ann Intern Med*. 2008;149(6):391-398.
7. Herzig SJ, Howell MD, Ngo LH, Marcantonio ER. Acid-suppressive medication use and the risk for hospital-acquired pneumonia. *JAMA*. 2009;301(20):2120-2128.
8. Dial S, Alrasadi K, Manoukian C, Huang A, Menzies D. Risk of *Clostridium difficile* diarrhea among hospital inpatients prescribed proton pump inhibitors: cohort and case-control studies. *CMAJ*. 2004;171(1):33-38.
9. Dial S, Delaney JA, Barkun AN, Suissa S. Use of gastric acid–suppressive agents and the risk of community-acquired *Clostridium difficile*–associated disease. *JAMA*. 2005;294(23):2989-2995.
10. Aseeri M, Schroeder T, Kramer J, Zackula R. Gastric acid suppression by proton pump inhibitors as a risk factor for *Clostridium difficile*–associated diarrhea in hospitalized patients. *Am J Gastroenterol*. 2008;103(9):2308-2313.
11. Cunningham R, Dale B, Undy B, Gaunt N. Proton pump inhibitors as a risk factor for *Clostridium difficile* diarrhoea. *J Hosp Infect*. 2003;54(3):243-245.
12. Nair S, Yadav D, Corpuz M, Pitchumoni CS. *Clostridium difficile* colitis: factors influencing treatment failure and relapse: a prospective evaluation. *Am J Gastroenterol*. 1998;93(10):1873-1876.
13. Tal S, Gurevich A, Guller V, Gurevich I, Berger D, Levi S. Risk factors for recurrence of *Clostridium difficile*–associated diarrhea in the elderly. *Scand J Infect Dis*. 2002;34(8):594-597.
14. Garey KW, Sethi S, Yadav Y, DuPont HL. Meta-analysis to assess risk factors for recurrent *Clostridium difficile* infection. *J Hosp Infect*. 2008;70(4):298-304.
15. Maroo S, Lamont JT. Recurrent *clostridium difficile. Gastroenterology*. 2006;130 (4):1311-1316.
16. Kelly CP, LaMont JT. *Clostridium difficile:* more difficult than ever. *N Engl J Med*. 2008;359(18):1932-1940.
17. Loo VG, Libman MD, Miller MA, et al. *Clostridium difficile:* a formidable foe. *CMAJ*. 2004;171(1):47-48.
18. Dubberke ER, Reske KA, Olsen MA, et al. Evaluation of *Clostridium difficile*–associated disease pressure as a risk factor for *C difficile*–associated disease. *Arch Intern Med*. 2007;167(10):1092-1097.
19. Dial S, Delaney JA, Schneider V, Suissa S. Proton pump inhibitor use and risk of community-acquired *Clostridium difficile*–associated disease defined by prescription for oral vancomycin therapy. *CMAJ*. 2006;175(7):745-748.
20. Dial S, Kezouh A, Dascal A, Barkun A, Suissa S. Patterns of antibiotic use and risk of hospital admission because of *Clostridium difficile* infection. *CMAJ*. 2008; 179(8):767-772.
21. Lowe DO, Mamdani MM, Kopp A, Low DE, Juurlink DN. Proton pump inhibitors and hospitalization for *Clostridium difficile*–associated disease: a population-based study. *Clin Infect Dis*. 2006;43(10):1272-1276.
22. Wilcox MH, Mooney L, Bendall R, Settle CD, Fawley WN. A case-control study of community-associated *Clostridium difficile* infection. *J Antimicrob Chemother*. 2008;62(2):388-396.
23. Kyne L, Warny M, Qamar A, Kelly CP. Association between antibody response to toxin A and protection against recurrent *Clostridium difficile* diarrhoea. *Lancet*. 2001;357(9251):189-193.
24. Cadle RM, Mansouri MD, Logan N, Kudva DR, Musher DM. Association of proton-pump inhibitors with outcomes in *Clostridium difficile* colitis. *Am J Health Syst Pharm*. 2007;64(22):2359-2363.
25. Teasley DG, Gerding DN, Olson MM, et al. Prospective randomised trial of metronidazole versus vancomycin for *Clostridium-difficile*–associated diarrhoea and colitis. *Lancet*. 1983;2(8358):1043-1046.
26. Zar FA, Bakkanagari SR, Moorthi KM, Davis MB. A comparison of vancomycin and metronidazole for the treatment of *Clostridium difficile*–associated diarrhea, stratified by disease severity. *Clin Infect Dis*. 2007;45(3):302-307.
27. Lai KK, Melvin ZS, Menard MJ, Kotilainen HR, Baker S. *Clostridium difficile*–associated diarrhea: epidemiology, risk factors, and infection control. *Infect Control Hosp Epidemiol*. 1997;18(9):628-632.
28. Baxter R, Ray GT, Fireman BH. Case-control study of antibiotic use and subsequent *Clostridium difficile*–associated diarrhea in hospitalized patients. *Infect Control Hosp Epidemiol*. 2008;29(1):44-50.
29. McFarland LV, Mulligan ME, Kwok RY, Stamm WE. Nosocomial acquisition of *Clostridium difficile* infection. *N Engl J Med*. 1989;320(4):204-210.
30. Bignardi GE. Risk factors for *Clostridium difficile* infection. *J Hosp Infect*. 1998; 40(1):1-15.

31. Moshkowitz M, Ben-Baruch E, Kline Z, Shimoni Z, Niven M, Konikoff F. Risk factors for severity and relapse of pseudomembranous colitis in an elderly population. *Colorectal Dis*. 2007;9(2):173-177.
32. Giannella RA, Broitman SA, Zamcheck N. Influence of gastric acidity on bacterial and parasitic enteric infections: a perspective. *Ann Intern Med*. 1973;78 (2):271-276.
33. Thorens J, Froehlich F, Schwizer W, et al. Bacterial overgrowth during treatment with omeprazole compared with cimetidine: a prospective randomised double blind study. *Gut*. 1996;39(1):54-59.
34. Gurian L, Ward TT, Katon RM. Possible foodborne transmission in a case of pseudomembranous colitis due to *Clostridium difficile:* influence of gastrointestinal secretions on *Clostridium difficile* infection. *Gastroenterology*. 1982;83(2): 465-469.
35. Jump RL, Pultz MJ, Donskey CJ. Vegetative *Clostridium difficile* survives in room air on moist surfaces and in gastric contents with reduced acidity: a potential mechanism to explain the association between proton pump inhibitors and *C difficile*–associated diarrhea? *Antimicrob Agents Chemother*. 2007;51(8):2883-2887.
36. Alarcón T, Domingo D, Sánchez I, Sanz JC, Martínez MJ, López-Brea M. In vitro activity of ebrotidine, ranitidine, omeprazole, lansoprazole, and bismuth citrate against clinical isolates of *Helicobacter pylori. Eur J Clin Microbiol Infect Dis*. 1998;17(4):275-277.
37. Zedtwitz-Liebenstein K, Wenisch C, Patruta S, Parschalk B, Daxbock F, Graninger W. Omeprazole treatment diminishes intra- and extracellular neutrophil reactive oxygen production and bactericidal activity. *Crit Care Med*. 2002;30(5): 1118-1122.
38. Yoshida N, Yoshikawa T, Tanaka Y, et al. A new mechanism for anti-inflammatory actions of proton pump inhibitors: inhibitory effects on neutrophil-endothelial cell interactions. *Aliment Pharmacol Ther*. 2000;14(suppl 1):74-81.
39. Agastya G, West BC, Callahan JM. Omeprazole inhibits phagocytosis and acidification of phagolysosomes of normal human neutrophils in vitro. *Immunopharmacol Immunotoxicol*. 2000;22(2):357-372.
40. Dubberke ER, Reske KA, Yan Y, Olsen MA, McDonald LC, Fraser VJ. *Clostridium difficile*–associated disease in a setting of endemicity: identification of novel risk factors. *Clin Infect Dis*. 2007;45(12):1543-1549.
41. Pham CQ, Regal RE, Bostwick TR, Knauf KS. Acid suppressive therapy use on an inpatient internal medicine service. *Ann Pharmacother*. 2006;40(7-8):1261-1266.
42. Grube RR, May DB. Stress ulcer prophylaxis in hospitalized patients not in intensive care units. *Am J Health Syst Pharm*. 2007;64(13):1396-1400.
43. Nardino RJ, Vender RJ, Herbert PN. Overuse of acid-suppressive therapy in hospitalized patients. *Am J Gastroenterol*. 2000;95(11):3118-3122.
44. 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(4):347-379.
45. Choudhry MN, Soran H, Ziglam HM. Overuse and inappropriate prescribing of proton pump inhibitors in patients with *Clostridium difficile*–associated disease. *QJM*. 2008;101(6):445-448.

Apixaban for Reduction In Stroke and Other ThromboembolLic Events in Atrial Fibrillation (ARISTOTLE) trial: Design and rationale

Renato D. Lopes, MD, PhD,[a] John H. Alexander, MD, MHS,[a] Sana M. Al-Khatib, MD, MHS,[a] Jack Ansell, MD,[b] Raphael Diaz, MD,[c] J. Donald Easton, MD,[d] Bernard J. Gersh, MB, ChB, DPhil,[e] Christopher B. Granger, MD,[a] Michael Hanna, MD,[f] John Horowitz, MD,[g] Elaine M. Hylek, MD, MPH,[h] John J. V. McMurray, MD,[i] Freek W. A. Verheugt, MD, PhD,[j] and Lars Wallentin, MD, PhD[k] on behalf of the ARISTOTLE Investigators

Durham, NC; New York, NY; Santa Fe, Argentina; Providence, RI; Rochester, MN; Princeton, NJ; Adelaide, Australia; Boston, MA; Glasgow, United Kingdom; Nijmegen, The Netherlands; and Uppsala, Sweden

Atrial fibrillation (AF) is associated with increased risk of stroke that can be attenuated with vitamin K antagonists (VKAs). Vitamin K antagonist use is limited, in part, by the high incidence of complications when patients' international normalized ratios (INRs) deviate from the target range. The primary objective of ARISTOTLE is to determine if the factor Xa inhibitor, apixaban, is noninferior to warfarin at reducing the combined endpoint of stroke (ischemic or hemorrhagic) and systemic embolism in patients with AF and at least 1 additional risk factor for stroke. We have randomized 18,206 patients from over 1,000 centers in 40 countries. Patients were randomly assigned in a 1:1 ratio to receive apixaban or warfarin using a double-blind, double-dummy design. International normalized ratios are monitored and warfarin (or placebo) is adjusted aiming for a target INR range of 2 to 3 using a blinded, encrypted point-of-care device. Minimum treatment is 12 months, and maximum expected exposure is 4 years. Time to accrual of at least 448 primary efficacy events will determine treatment duration. The key secondary objectives are to determine if apixaban is superior to warfarin for the combined endpoint of stroke (ischemic or hemorrhagic) and systemic embolism, and for all-cause death. These will be tested after the primary objective using a closed test procedure. The noninferiority boundary is 1.38; apixaban will be declared noninferior if the 95% CI excludes the possibility that the primary outcome rate with apixaban is >1.38 times higher than with warfarin. ARISTOTLE will determine whether apixaban is noninferior or superior to warfarin in preventing stroke and systemic embolism; whether apixaban has particular benefits in the warfarin-naïve population; whether it reduces the combined rate of stroke, systemic embolism, and death; and whether it impacts bleeding. (Am Heart J 2010;159:331-9.)

Atrial fibrillation (AF) is a growing public health problem worldwide and is the most common arrhythmia requiring hospitalization in the United States.[1] The incidence of AF has increased over the past 2 to 3 decades. The number of people with AF in the United States is projected to exceed 10 million by 2050[2]; this will be accompanied by substantial morbidity and mortality.[2,3] Stroke is considered to be the most significant morbidity in patients with AF. Approximately 15% of strokes occur in those with AF and the risk of stroke in untreated AF patients averages 5% per year.[3,4] The annual risk of stroke in AF patients is related to age, increasing from 1.5% for patients aged 50 to 59 years to 23% for those aged 80 to 89 years.[5] When compared with stroke from other causes, ischemic stroke secondary to AF carries twice the risk of death.[5]

Warfarin

Warfarin, a vitamin K antagonist (VKA), reduces the risk of stroke by approximately 62%[6]; however, VKAs have major limitations and are underused in clinical practice.[7] Patients are frequently outside the optimal target anticoagulation range when VKAs are prescribed, exposing them to the risk of thrombosis or bleeding.[8,9] Clinical trial data show that among patients receiving

From the [a]Duke Clinical Research Institute, Duke University Medical Center, Durham, NC, [b]Lenox Hill Hospital, New York, NY, [c]Estudios Cardiologicos Latinoamerica, Santa Fe, Argentina, [d]Rhode Island Hospital, Alpert Medical School, Brown University, Providence, RI, [e]Mayo Clinic College of Medicine, Rochester, MN, [f]Bristol-Myers Squibb, Princeton, NJ, [g]University of Adelaide, Adelaide, Australia, [h]Boston University Medical Center, Boston, MA, [i]Western Infirmary, Glasgow, United Kingdom, [j]University Medical Center, Nijmegen, The Netherlands, and [k]Uppsala Clinical Research Center, Uppsala University, Uppsala, Sweden.

RCT reg no. NCT00412984.

Submitted February 6, 2009; accepted July 7, 2009.

Reprint requests: Christopher B. Granger, MD, Duke Clinical Research Institute, Duke University Medical Center, 2400 Pratt Street, Room 0311 Terrace Level, Durham, NC 27707.

E-mail: christopher.granger@duke.edu

0002-8703/$ - see front matter

doi:10.1016/j.ahj.2009.07.035

warfarin or other VKAs, even in the highly structured setting of a trial, the international normalized ratio (INR) is outside the therapeutic range (2.0-3.0) about one third of the time.[10-13] Moreover, a meta-analysis shows that the proportion of time outside the therapeutic INR range in community practice is 43%.[11] Thus, a large number of patients with AF who should be on warfarin are either not deriving full benefit from warfarin or are not receiving it at all. Not only do VKAs have a narrow therapeutic window but they also exhibit a highly variable dose response that is attributable to genetic, disease-related, and environmental factors. Their dosing can be particularly challenging among elderly individuals due to changes in the pharmacokinetics and pharmacodynamics that occur with age. Other factors influencing drug effect include prescription and nonprescription drugs, dietary vitamin K, and botanical products.[14,15] The need for regular and lifelong therapeutic monitoring is an inconvenience for many patients. These limitations of VKAs illustrate the need for new oral anticoagulants for the prevention of thromboembolism in patients with AF.

Current guidelines, including those from the American College of Cardiology/American Heart Association/European Society of Cardiology, recommend a risk-based approach for antithrombotic therapy for stroke prevention.[16] Aspirin or VKAs are recommended for patients with AF who have one moderate risk factor for stroke ($CHADS_2$ risk score of 1), and VKAs are recommended for patients with more than one moderate risk factor ($CHADS_2$ risk score ≥ 2) for stroke.[16] Vitamin K antagonists are efficacious at preventing stroke even in the lower risk AF population[17]; however, their routine use is not currently recommended because the overall net benefit in this low-risk group is uncertain. Because the oral factor Xa inhibitors avoid many of the limitations of VKAs, it is important to evaluate these new agents across the spectrum of risk including patients with a low to intermediate risk of stroke.

Another important issue is that in patients starting warfarin for the first time, the risk of thromboembolism and bleeding is high for the first year due in part to the time taken to establish an adequate and stable INR. In one observational study, patients starting warfarin had a 3-fold increased risk of bleeding in the first 90 days of treatment[7] compared with patients already on warfarin.[7] The ACTIVE-W study showed that, in patients with AF at high risk of stroke, warfarin was superior to clopidogrel plus aspirin for prevention of vascular events.[18] Additional analyses also showed that the rates of discontinuation of warfarin therapy during the first year of the trial were higher (15%) in the warfarin-naïve patients when compared with the warfarin-experienced patients (8%).[19]

In antithrombotic therapy for the prevention of thromboembolism in patients with AF, there is a need for an alternative to warfarin that has greater efficacy, greater convenience of use, and/or greater safety. When compared with warfarin in the SPORTIF III and V trials, the oral direct thrombin inhibitor ximelagatran was found to be effective with an acceptable bleeding risk,[10,12] but its development was stopped due to liver toxicity. A number of new anticoagulants, including factor Xa inhibitors and direct thrombin inhibitors, are now being developed and evaluated as alternatives to warfarin for stroke prevention in AF. The ideal agent should be oral and free from variation in absorption related to food intake and should have predictable pharmacokinetics, few drug interactions, and minimal toxicity. It would not require therapeutic monitoring and its efficacy and safety would be comparable with or superior to warfarin at preventing thromboembolism in AF subjects.

Apixaban

Factor Xa occupies a pivotal role in the clotting cascade because it promotes conversion of prothrombin to thrombin. Developed as an anticoagulant agent, apixaban is an orally active selective inhibitor of the coagulation factor Xa. The direct mechanism of this drug does not require the presence of antithrombin. It has a predominantly nonrenal (75%) clearance and a half-life around 12 hours in healthy volunteers. The effect is independent of vitamin K intake, and with the exception of strong CYP3A4 inhibitors, there is minimal potential for drug-to-drug interaction. This compound has shown efficacy in preclinical animal models for venous and arterial thrombosis and has no organ-specific toxicity in animal models of up to 12 months of exposure.[20,21]

In a phase II dose-ranging study of deep vein thrombosis (DVT) prevention in patients undergoing knee replacement surgery, pooling of all doses of apixaban showed a 21% reduction in venous thromboembolism (VTE) or all-cause death when compared with enoxaparin and a 53% reduction in VTE or all-cause death when compared with warfarin.[22] Importantly, the frequency of major bleeding events was low (0%-3.3%) and comparable among all apixaban arms. A dose-ranging trial in DVT treatment including 520 patients with proximal DVT also showed favorable efficacy and safety, including the 5 mg twice a day apixaban dose arm.[23]

The ADVANCE-1 study, a phase III VTE prevention trial, compared apixaban, 2.5 mg, twice daily with the Food and Drug Administration-approved dose of enoxaparin (30 mg twice daily).[24] The primary efficacy outcome was a composite of symptomatic or asymptomatic DVT, pulmonary embolism, and death from any cause. In a preliminary analysis, the rate of the primary efficacy endpoint in the apixaban arm was similar to that observed with enoxaparin (9.0% vs 8.9%, $P = .64$). The major bleeding rate for apixaban, however, tended to be

lower than for enoxaparin (0.7% vs 1.4%, $P = .053$), and the composite rate of clinically relevant nonmajor bleeding and major bleeding was significantly lower in patients assigned to apixaban than those assigned to enoxaparin (2.9% vs 4.3%, $P = .034$).

APPRAISE was a phase II, randomized, double-blind, placebo-controlled, parallel arm study that enrolled 1,715 patients with acute coronary syndrome (ACS) from Europe and North America.[25] This study demonstrated that the addition of apixaban to current antiplatelet therapy for 6 months after ST elevation or non-ST elevation ACS results in a dose-dependent increase in bleeding and a trend toward a reduction in ischemic events. Based on these data, apixaban at a total daily dose of between 5 and 10 mg is being tested in patients with ACS receiving aspirin or dual antiplatelet therapy in the phase III trial APPRAISE 2.

These findings indicate that oral factor Xa inhibition at a dose of 5 to 10 mg daily may be an effective and safe approach toward anticoagulation. Based on the experience in DVT prevention and treatment trials, a dose of 5 mg twice a day of apixaban was chosen for comparison with warfarin for prevention of stroke and systemic embolism in patients with AF in the ARISTOTLE trial.

Study objectives

Primary objective

The primary objective of this study is to determine whether apixaban is noninferior to warfarin (INR target range 2.0-3.0) at reducing the combined outcome of stroke (ischemic or hemorrhagic) and systemic embolism in subjects with AF and at least 1 additional risk factor for stroke.

Secondary objectives

The key secondary objectives are to determine if apixaban is superior to warfarin for the combined endpoint of stroke (ischemic or hemorrhagic) and systemic embolism, and for all-cause death. These will be tested after the primary objective using a closed test procedure.

Other secondary objectives are to compare apixaban and warfarin with respect to:

- The composite endpoint of ischemic stroke, hemorrhagic stroke, systemic embolism, and all-cause death,
- The composite endpoint (in warfarin-naïve patients) of ischemic stroke, hemorrhagic stroke, systemic embolism and major bleeding,
- The composite endpoint of ischemic stroke, hemorrhagic stroke, systemic embolism, and major bleeding,
- The composite endpoint of ischemic stroke, hemorrhagic stroke, systemic embolism, major bleeding, and all-cause death,
- The composite endpoint of ischemic stroke, hemorrhagic stroke, systemic embolism, myocardial infarction, and all-cause death,
- Major bleeding

Table I. Inclusion and exclusion criteria

Inclusion criteria

Age ≥18 y
Permanent or persistent AF or atrial flutter on ECG at enrollment; or AF or atrial flutter documented by ECG or as an episode ≥1 min on rhythm strip, Holter monitor, or intracardiac recording on 2 separate occasions at least 2 wk apart in 12 mo before enrollment
One or more of the following risk factors for stroke
  Age ≥75 y
  Prior stroke, TIA, or systemic embolus
  Symptomatic CHF within 3 mo or LV dysfunction with LVEF ≤40% by echocardiography, radionuclide study, or contrast angiography
  Diabetes mellitus
  Hypertension requiring pharmacologic treatment
Women of childbearing potential must use contraception to avoid pregnancy during treatment period or for 2 wk after last dose of study medication, whichever is longer
All subjects must provide signed written informed consent

Exclusion criteria

AF or atrial flutter due to reversible causes (eg, thyrotoxicosis, pericarditis)
Clinically significant (moderate or severe) mitral stenosis
Increased bleeding risk believed to be a contraindication to oral anticoagulation (eg, previous intracranial hemorrhage)
Conditions other than AF that require chronic anticoagulation (eg, prosthetic mechanical heart valve)
Persistent uncontrolled hypertension (SBP >180 mm Hg or DBP >100 mm Hg)
Active infective endocarditis
Planned major surgery
Planned AF or atrial flutter ablation procedure
Use of unapproved investigational drug or device in past 30 d
Required aspirin >165 mg/d
Simultaneous treatment with both aspirin and a thienopyridine (eg, clopidogrel, ticlopidine)
Severe comorbid condition with life expectancy ≤1 y
Active alcohol or drug abuse or psychosocial reasons that make study participation impractical
Recent stroke (within 7 d)
Severe renal insufficiency (serum creatinine level >2.5 mg/dL or calculated creatinine clearance <25 mL/min)
ALT or AST >2 × ULN or a total bilirubin ≥1.5 × ULN (unless an alternative causative factor [eg, Gilbert's syndrome] is identified)
Platelet count ≤100,000/mm^3
Hemoglobin level <9 g/dL
Inability to comply with INR monitoring

ECG, Electrocardiogram; *CHF*, congestive heart failure; *LV*, left ventricular; *LVEF*, left ventricular ejection fraction; *SBP*, systolic blood pressure; *DBP*, diastolic blood pressure; ALT, alanine transaminase; AST, aspartate transaminase; *ULN*, upper limit of normal; *wk*, weeks; *mo*, months.

Study population

In this double-blind study, we have randomized 18,206 patients with AF from over 1,000 centers in about 40 countries. Eligible subjects were randomly assigned in a 1:1 ratio to receive either apixaban or warfarin titrated to a target INR range of 2.0 to 3.0. Both warfarin-naïve and warfarin-experienced patients are being recruited, with a

special emphasis on the former. Our aim is for 40% of patients to be warfarin naïve. The inclusion and exclusion criteria are summarized in Table I.

Randomization and study drug

Subjects who were on warfarin before randomization discontinued the drug 72 hours before randomization and were not dosed with study drug until the INR was <2.0. Randomization is stratified by investigative site and prior warfarin use status (experienced or naïve). Patients are classified as warfarin naïve if they have never used warfarin (or other VKAs) or if they have used it for ≤30 consecutive days. Otherwise, patients are considered warfarin experienced.

To maintain blinding, study medications are packaged using a double-dummy design. The 2 sets of tablets each subject receives are distinguishable by color and size, but active apixaban tablets match placebo apixaban tablets and active warfarin tablets match placebo warfarin tablets to ensure blinding of the patient and investigator. After randomization, patients receive either apixaban and warfarin placebo or apixaban placebo and warfarin. During the titration phase, we recommend the use of a dosing algorithm with initial daily dose of up to 6 mg of warfarin (or warfarin placebo) (unless already on a stable dose of warfarin, in which case that may be continued) and dose of apixaban (or apixaban placebo) of 5 mg twice a day. For patients who are estimated to have higher apixaban drug exposure, we will use a lower dose of 2.5 mg twice a day of apixaban. Subjects who fulfill any 2 of the following criteria at baseline will receive the lower apixaban dose of 2.5 mg twice a day: age ≥80 years, body weight ≤60 kg, and serum creatinine level ≥1.5 mg/dL (133 μmol/L). Subsequent warfarin doses are recommended based upon an algorithm provided to the investigators; however, the final dose decision is left to the discretion of the investigator. Subjects, investigators, members of the steering and adjudication committees, and the sponsor's staff conducting the study do not have access to individual subject treatment assignments.

Study design and duration

The trial is event driven, thus, the number of subjects required and length of treatment were estimated based on event rates in similar trials. The final duration of the trial will be determined by the time required to accrue at least 448 primary efficacy events (see statistical methods). All subjects will be followed from randomization until the study end date.

The study includes a screening period of up to 14 days. Subjects with AF and at least ≥1 risk factors for stroke will be evaluated for study eligibility. The design of the study is shown in Figure 1.

INR monitoring

The INR monitoring for the warfarin-naïve patients begins on the fourth day after initiation of study drug and is performed twice a week for 2 weeks, once a week for 2 weeks, and monthly thereafter once a stable INR is attained. For the warfarin-experienced patients who have been on stable dosing of VKA for at least 3 months, INR monitoring visits are required on day 1, week 1, week 2, and then monthly. An investigator may increase the frequency of INR monitoring if it is considered clinically indicated.

Titration of the study drug is based on central monitoring of INR measurements using encrypted point-of-care (POC) devices, centralized dosing recommendations, and sham apixaban titration. The POC device delivers an encrypted result to the investigator who telephones or electronically transmits the result along with the subject's identification number, date, and time to a central response system. This system processes the information in a blinded manner and returns either a true INR value (if a subject is receiving warfarin) or a sham INR value (if a subject is receiving apixaban). Although investigators will need to obtain open label INR values when clinically indicated, efforts will be made to minimize nonstudy INR assessments. The final dosing decision rests with the investigator. Assessments of outcomes and study medication compliance are performed at each INR visit.

Follow-up

The follow-up period will last until the attainment of at least 448 primary study events. Follow-up of subjects who discontinue study drug will occur quarterly until the end of the study.

Outcome definitions

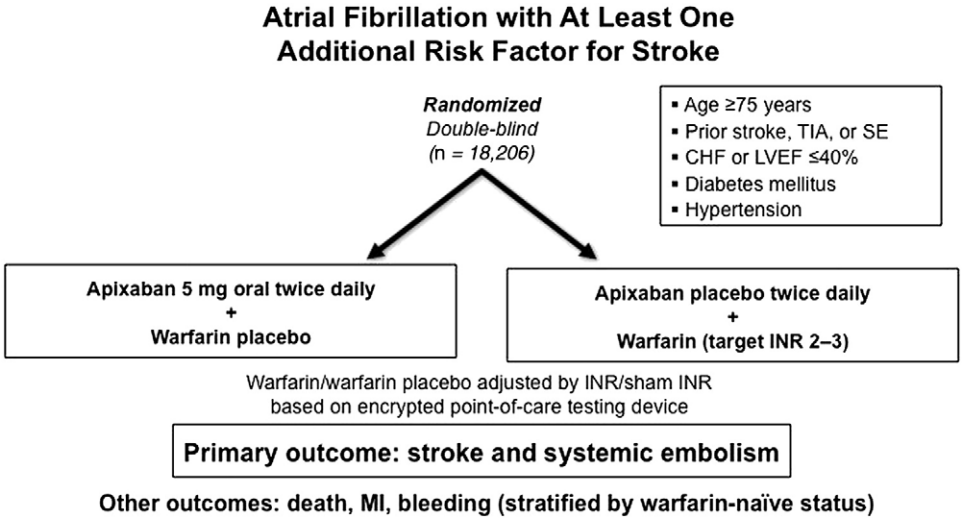
Efficacy outcomes

The primary efficacy outcome is the time to first occurrence of stroke (ischemic or hemorrhagic) or systemic embolism.

In this study, stroke is defined as a nontraumatic abrupt onset of a focal neurologic deficit lasting at least 24 hours. A retinal ischemic event (embolism or thrombosis) will be considered a stroke. A cerebral imaging study (computed tomographic scan or magnetic resonance imaging) is recommended for all suspected strokes. Strokes will be classified as ischemic, ischemic with hemorrhagic transformation, hemorrhagic, or of uncertain type. Hemorrhagic strokes will be subclassified as subdural, subarachnoid, or intraparenchymal.

A transient ischemic attack (TIA) is defined as a nontraumatic abrupt onset of a focal neurologic deficit lasting <24 hours. Stroke and TIA will be further subclassified based on whether there is imaging evidence of a new cerebral infarction that correlates with the clinical presentation of the subject.

Figure 1



Study design.

The diagnosis of systemic embolism requires a clinical history consistent with an acute loss of blood flow to a peripheral artery (or arteries) supported by evidence of embolism from surgical specimens, autopsy, angiography, vascular imaging, or other objective testing.

Safety outcomes

The primary safety endpoint is time to first occurrence of confirmed major bleeding.

The definition of major bleeding described below is adapted from the protocol and the International Society on Thrombosis and Hemostasis definition. The baseline hemoglobin level is defined as the closest hemoglobin level value before the bleeding event.

Major bleeding is defined as acute or subacute clinically overt bleeding accompanied by ≥1 of the following: (1) a decrease in hemoglobin level of ≥2 g/dL over a 24-hour period; (2) a transfusion of ≥2 U of packed red blood cells; and/or (3) bleeding that is fatal or occurs in at least 1 of the following critical sites: intracranial, intraspinal, intraocular, pericardial, intraarticular, intramuscular with compartment syndrome, retroperitoneal.

Clinically relevant nonmajor bleeding is defined as acute or subacute clinically overt bleeding that does not satisfy the criteria for major bleeding and leads to hospital admission for bleeding, physician-guided medical or surgical treatment for bleeding, or a change in antithrombotic therapy (including study drug) for bleeding.

All acute clinically overt bleeding events not meeting the criteria for either major bleeding or clinically relevant nonmajor bleeding are classified as minor bleeding. Bleeding events are also classified by the TIMI[26] and GUSTO criteria.[27]

The secondary safety outcome for this trial is a composite of major bleeding and clinically relevant nonmajor bleeding. Other safety outcome measures include minor bleeding, fractures, and other adverse events.

Clinical Events Committee adjudication

An independent, blinded, clinical events committee (CEC) adjudicates all suspected hemorrhagic and nonhemorrhagic strokes, TIAs, systemic emboli, major and clinically relevant non-major bleeding, myocardial infarction, and cause of death.

Using prespecified event definitions and agreed upon event adjudication criteria, the CEC adjudicates suspected events based on the preponderance of the evidence and the clinical knowledge and experience of the physician reviewers. Event adjudication in ARISTOTLE occurs in 2 phases. All suspected events are adjudicated in phase I. Each stroke, systemic embolism, and bleeding event are reviewed by 2 independent physician reviewers. Each myocardial infarction and death event are reviewed by 1 reviewer. Significant disagreements between phase I reviewers for stroke, systemic embolism, and bleeding cases are identified as needing phase II review. In addition, a quality control sample undergoes phase II review. All phase II reviews are conducted by committee with the final adjudicated result being a consensus of the committee members present. In both phase I and phase II reviews, all stroke events are evaluated by at least 1 neurologist.

Statistical analysis and sample size calculation

A meta-analysis of 6 AF studies estimated a relative reduction in the risk of stroke or systemic thromboembolism of 62% for warfarin versus placebo.[6] Based on historical trials, the primary noninferiority hypothesis of ARISTOTLE is that apixaban will preserve at least 50% of the benefit of warfarin in preventing stroke and systemic embolism. This gives an upper CI of 1.88 of the apixaban versus warfarin relative risk (Figure 2*A*). To be more certain that apixaban is noninferior to warfarin based on this single clinical trial, more stringent boundaries are defined. In response to different international regulatory bodies, 2 noninferiority tests will be applied. First, the 95% CI should not include ≥1.38 to declare noninferiority (Figure 2*B*). In addition, the 99% CI should not include ≥1.44 to declare noninferiority. Because the event rate is lower than initially expected, it is estimated that approximately 18,000 randomized patients with sufficient risk of stroke and sufficient treatment duration will result in 448 patients with primary outcome events needed for 90% power to meet the primary objective of the study. We originally estimated that the follow-up would be an average of 1.8 years, assuming a stroke or systemic embolism rate of 1.67 per 100 subject-years.

An independent Data Monitoring Committee is charged with monitoring the accumulating trial data. A formal interim analysis will be performed once 50% of the primary efficacy endpoint events have been confirmed by the CEC. The objective of this interim analysis is to determine whether apixaban is superior to warfarin for the primary efficacy endpoint. No interim testing for noninferiority will be performed.

Pharmacokinetic biomarkers

The main objectives of the biomarker and genetic substudy program are to correlate genetic polymorphisms and levels of biomarkers with clinical outcomes, to improve risk stratification for stroke among patients with AF, and to relate the effects of apixaban and warfarin to these disease biomarkers. Several biomarkers will be analyzed at baseline for as many of the 18,206 patients as possible (Table II), and a second blood sample will be collected for approximately 5,000 patients at 2 months. This will allow the analysis of changes in biomarker levels.

Organizational structure

The ARISTOTLE trial is led by an academic steering committee composed of 2 cochairs, national coordinators from each participating country, and a representative from the trial sponsor. This committee provides scientific direction and input, addresses policy issues regarding the protocol, and meets periodically to assess the trial progress.

The ARISTOTLE executive committee, composed of a subset of senior leaders from the steering committee, is responsible for evaluating the progress and safety of the trial and making decisions regarding early termination or continuation of the trial.

A subset of the steering committee will form the publications committee that will oversee the publication

Figure 2

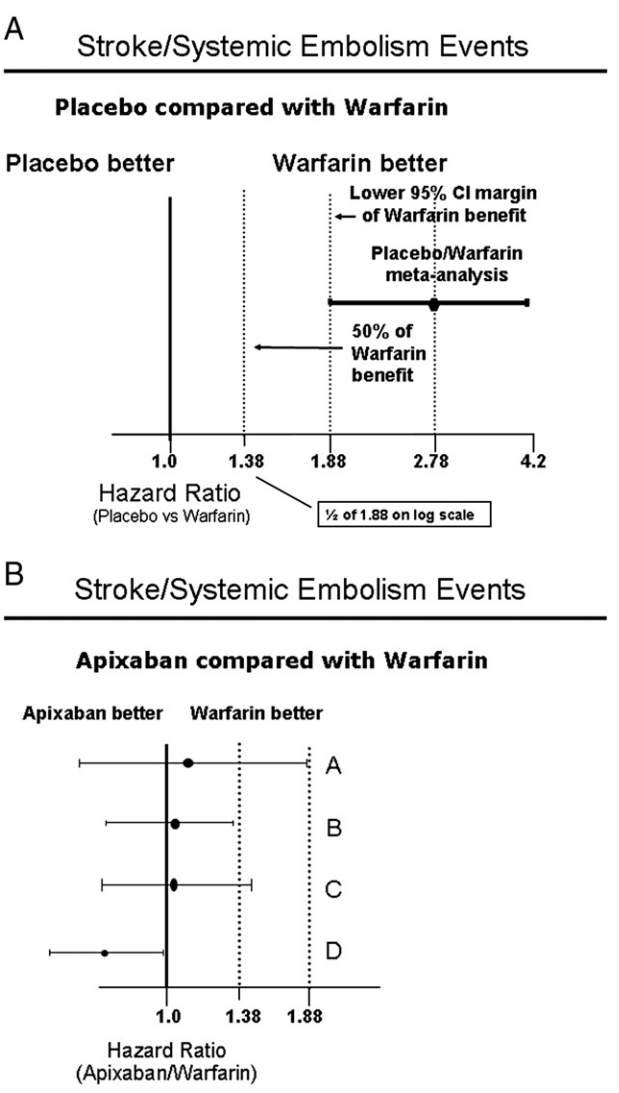


Illustration and interpretation of noninferiority boundaries for warfarin comparator AF trials. **A**, Noninferiority boundaries for warfarin comparator AF trials. **B**, Apixaban preserves some benefit (A), apixaban preserves >50% of warfarin benefit (B), apixaban preserves more than about 40% of warfarin benefit (C), and apixaban is superior to warfarin (D).

Table II. Biomarkers

Biomarkers	Area	Characteristic
hs-Troponin I	Myocardial necrosis, myocardial function	Strong predictor for raised mortality in ACS and in healthy elderly
NT-proBNP	Myocardial necrosis, myocardial function	↑ in patients with AF; high level indicates increased risk for thromboembolism
ADMA	Endothelial function	↑ in patients with vascular disease
vWF	Endothelial function	↑ levels in AF patients; predictor for vascular events
hs-CRP, IL-6	Inflammation	Associated with AF and risk of future CV events in healthy individuals
Soluble CD40 ligand	Platelet activity	↑ levels related to inflammatory activity, coagulation, and platelet activation
Fragment 1+2	Coagulation	↑ in patients with AF and related to risk factors for stroke; reduced with warfarin treatment
D-dimer	Coagulation	↑ in patients with AF and associated with new thrombotic events; reduced with warfarin treatment
Cystatine C	Renal function	↑ in patients with reduced renal function; poorer prognosis in patients with CV disease; better marker for endogenous GFR than creatinine clearance
HbA1C, Apo A, Apo B	Metabolism and lipoproteins	Lipoproteins and diabetes mellitus risk factors for CV disease; diabetes mellitus predictor of an increased risk for complications

BNP, Brain natriuretic peptide; *CV*, cardiovascular; *GFR*, glomerular filtration rate.

process for the main manuscript and all secondary presentations and manuscripts resulting from the trial.

The ARISTOTLE Trial is sponsored and funded by Bristol-Myers Squibb (Princeton, NJ, USA) and Pfizer (New York, NY, USA).

Discussion

Atrial fibrillation is a very common arrhythmia and is associated with an increased risk of mortality and morbidity, particularly stroke, which is the third most common cause of death in developed countries and the leading cause of serious long-term disability worldwide.[28]

The ACTIVE-W study included 6,706 patients with AF and at least ≥1 risk factors for stroke.[18] ACTIVE-W was designed to compare oral anticoagulation therapy with aspirin plus clopidogrel for prevention of vascular events in patients with AF at high risk for stroke. This study was not blinded. Moreover, there was no requirement for a certain proportion of warfarin-naïve patients. At study entry, approximately 76% of the patients were on warfarin. The study was stopped early because of the superiority of warfarin over aspirin plus clopidogrel for prevention of vascular events in patients with AF. These results were driven by the higher rates of stroke on clopidogrel plus aspirin.

Several oral direct thrombin inhibitors are being tested for stroke prevention in patients with AF such as ximelagatran and dabigatran.[29] This class of drug has a wider therapeutic range than warfarin, low potential for food and drug interactions, and no requirements for dose adjustments or regular monitoring. The SPORTIF III and V trials involving 3,407 and 3,922 patients, respectively, showed no difference in stroke prevention between ximelagatran and warfarin.[10,12] However, there was a significant increase in the level of liver enzymes in patients on ximelagatran. Moreover, major adverse cardiovascular events were observed in other studies,[30,31] and the drug was withdrawn from all markets in 2006. In addition, there was an unexplained higher rate of major bleeding in SPORTIF V when compared with SPORTIF III. Recently the RE-LY open label trial reported the results comparing 110 mg twice daily and 150 mg twice daily of the oral thrombin inhibitor dabigatran versus warfarin in 18,113 AF paitents with at least one risk factor for stroke. In this study dabigatran 150 mg twice daily was superior and 110 mg twice daily noninferior concerning stroke while simultaneously both doses reduced intracranial and life-threatening hemorrhage.[32] Therefore, there is now a larger focus on testing for potential superiority of apixaban over warfarin in the current trial.

Other phase III trials are testing inhibitors of factor Xa for stroke prevention in patients with AF. ROCKET-AF is comparing warfarin with rivaroxaban in a double-blind trial of 14,000 patients with AF at high risk for stroke.[33] AVERROES is another double-blind trial of apixaban in comparison with aspirin in 5,600 patients with AF and moderate risk of stroke or intolerance to warfarin.[34] BOREALIS-AF is testing subcutaneous weekly injections of biotinylated idraparinux, a subcutaneous indirect factor Xa inhibitor, in patients with AF and high risk of stroke.[35] This trial is designed as a double-blind comparison of idraparinux with warfarin in 9,600 patients.

ARISTOTLE will define whether apixaban is noninferior or/and superior to warfarin in the overall population and whether it may be superior solely in the warfarin-naïve population, which is more susceptible to complications until a stable warfarin dose has been established and more prone to warfarin discontinuation.

Important features of the ARISTOTLE trial include the investigation of a drug with a relatively long half-life and robust phase II to III data on safety and efficacy of the tested dose for DVT prophylaxis, the double-blind design with use of encrypted POC INR monitoring devices, good representation of warfarin-naïve patients, inclusion of the full risk-spectrum of patients ($CHADS_2$ score ≥1),

multinational representation of various health care systems and ethnic groups, and modification of dose of the study drug for patients with the highest drug exposure (2 criteria of older age, renal insufficiency, or low body weight). In ARISTOTLE, the protocol provides guidance on management of bleeding. Most bleeding can be managed by discontinuation of antithrombotic therapy (including study medication), local hemostatic measures, and fresh frozen plasma as needed. Patients with ongoing serious or life-threatening bleeding may require unblinding to guide reversal with vitamin K, and consideration of use of prothrombin complex concentrate, if on warfarin. For patients on apixaban, reversal of the anticoagulant effects will occur relatively rapidly over time given its half-life of around 12 hours. Apixaban has no specific antidote or reversal agent. Other agents, such as recombinant activated factor VII (NovoSeven, Novo Nordisk A/S, Bagsvaerd, Denmark), have not been studied in this setting and are not recommended.

Conclusion

In conclusion, the ARISTOTLE trial will answer many important questions related to stroke prevention in AF. Most importantly, it will compare apixaban with warfarin for stroke prevention in a wide range of patients with AF. It will provide information about the efficacy of apixaban both in warfarin-experienced and warfarin-naïve patients. Finally, it will generate a better understanding of this common disease and the risks of stroke and bleeding based on large scale substudies on genetic and protein biomarkers.

References

1. Wattigney WA, Mensah GA, Croft JB. Increasing trends in hospitalization for atrial fibrillation in the United States, 1985 through 1999: implications for primary prevention. Circulation 2003; 108:711-6.
2. 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.
3. Benjamin EJ, Wolf PA, D'Agostino RB, et al. Impact of atrial fibrillation on the risk of death: the Framingham Heart Study. Circulation 1998;98:946-52.
4. Wolf PA, D'Agostino RB, Belanger AJ, et al. Probability of stroke: a risk profile from the Framingham Study. Stroke 1991;22:312-8.
5. Lairikyengbam SK, Anderson MH, Davies AG. Present treatment options for atrial fibrillation. Postgrad Med J 2003;79:67-73.
6. Hart RG, Benavente O, McBride R, et al. Antithrombotic therapy to prevent stroke in patients with atrial fibrillation: a meta-analysis. Ann Intern Med 1999;131:492-501.
7. 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.
8. Goto S, Bhatt DL, Röther J, et al. Prevalence, clinical profile, and cardiovascular outcomes of atrial fibrillation patients with atherothrombosis. Am Heart J 2008;156:855-63.
9. Wan Y, Heneghan C, Perera R, et al. Anticoagulation control and prediction of adverse events in patients with atrial fibrillation: a systematic review. Circulation Cardiovascular Quality and Outcomes 2008;1:84-91.
10. 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.
11. van Walraven C, Jennings A, Oake N, et al. Effect of study setting on anticoagulation control: a systematic review and metaregression. Chest 2006;129:1155-66.
12. 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;362:1691-8.
13. 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: 546S-92S.
14. Holbrook AM, Pereira JA, Labiris R, et al. Systematic overview of warfarin and its drug and food interactions. Arch Intern Med 2005; 165:1095-106.
15. Cheng TO. Warfarin interaction with herbal drugs and food. Int J Cardiol 2007;119:107-8.
16. Fuster V, Rydén 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-e354.
17. 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;39:1482-6.
18. The ACTIVE Writing Group on behalf of the ACTIVE Investigators. 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.
19. 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;118:2029-37.
20. Wong PC, Crain EJ, Xin B, et al. Apixaban, an oral, direct and highly selective factor Xa inhibitor: in vitro, antithrombotic and antihemostatic studies. J Thromb Haemost 2008;6:820-9.
21. Pinto DJ, Orwat MJ, Quan ML, et al. 1-[3-Aminobenzisoxazol-5′-yl]-3-trifluoromethyl-6-[2′-(3-(*R*)-hydroxy-*N*-pyrrolidinyl)methyl-[1, 1′]-biphen-4-yl]-1,4,5,6-tetrahydropyrazolo-[3,4-c]-pyridin-7-one (BMS-740808) a highly potent, selective, efficacious, and orally bioavailable inhibitor of blood coagulation factor Xa. Bioorg Med Chem Lett 2006;16:4141-7.
22. Lassen MR, Davidson BL, Gallus A, et al. The efficacy and safety of apixaban, an oral, direct factor Xa inhibitor, as thromboprophylaxis in patients following total knee replacement. J Thromb Haemost 2007;5:2368-75.
23. Buller HR. A dose-finding study of the oral direct factor Xa inhibitor apixaban in the treatment of patients with acute symptomatic deep

vein thrombosis. Presented as a late-breaking clinical trial abstract at: 21st Congress of the International Society on Thrombosis and Haemostasis; July 9, 2007; Geneva, Switzerland; 2007.
24. Bristol-Myers Squibb and Pfizer provide update on apixaban clinical development program [press release]. Princeton, NJ & New York, NY: Business Wire; August 26, 2008.
25. Alexander JH, Becker RC, Bhatt DL, et al. Apixaban, an oral, direct, selective factor Xa inhibitor, in combination with antiplatelet therapy after acute coronary syndrome: results of the Apixaban for Prevention of Acute Ischemic and Safety Events (APPRAISE) trial. Circulation 2009;119:2877-85.
26. Bovill EG, Terrin ML, Stump DC, et al. Hemorrhagic events during therapy with recombinant tissue-type plasminogen activator, heparin, and aspirin for acute myocardial infarction. Results of the Thrombolysis in Myocardial Infarction (TIMI), Phase II Trial. Ann Intern Med 1991;115:256-65.
27. An international randomized trial comparing four thrombolytic strategies for acute myocardial infarction. The GUSTO Investigators. N Engl J Med 1993;329:673-82.
28. Lloyd-Jones D, Adams RJ, Brown TM, et al. Heart Disease and Stroke Statistics—2010 Update. A Report From the American Heart Association. Circulation 2009.
29. DiNisio M, Middeldorp S, Büller HR. Direct thrombin inhibitors. N Engl J Med 2005;353:1028-40.
30. Hermans C, Claeys D. Review of the rebound phenomenon in new anticoagulant treatments. Curr Med Res Opin 2006;22:471-81.
31. Fiessinger JN, Huisman MV, Davidson BL, et al. Ximelagatran vs low-molecular-weight heparin and warfarin for the treatment of deep vein thrombosis: a randomized trial. JAMA 2005;293:681-9.
32. Connolly SJ, Ezekowitz MD, Yusuf S, et al. Dabigatran versus warfarin in patients with atrial fibrillation. N Engl J Med 2009;361:1139-51.
33. Randomized, Double-Blind Study Comparing Once Daily Oral Rivaroxaban With Adjusted-Dose Oral Warfarin for the Prevention of Stroke in Subjects with Non-Valvular Atrial Fibrillation. ClinicalTrials.gov Identifier: NCT00403767. ClinicalTrials.gov Web site. http://clinicaltrials.gov/ct2/show/NCT00403767?term = nct00403767&rank=1. Last accessed February 4, 2009.
34. A Phase III Study of Apixaban in Patients With Atrial Fibrillation (AVERROES). ClinicalTrials.gov Identifier: NCT00496769. Clinical Trials.gov Web site. http://clinicaltrials.gov/ct2/show/NCT00496769?term=nct00496769&rank=1. Last accessed February 4, 2009.
35. Evaluation of Weekly Subcutaneous Biotinylated Idraparinux Versus Oral Adjusted-Dose Warfarin to Prevent Stroke and Systemic Thromboembolic Events in Patients with Atrial Fibrillation (BOREALIS-AF). ClinicalTrials.gov Identifier: NCT00580216. ClinicalTrials.gov Web site. http://clinicaltrials.gov/ct2/show/NCT00580216?term=nct00580216&rank=1. Last accessed February 4, 2009.

Social Science & Medicine 71 (2010) 187–195

Contents lists available at ScienceDirect

Social Science & Medicine

journal homepage: www.elsevier.com/locate/socscimed

Life course socioeconomic position is associated with inflammatory markers: The Framingham Offspring Study☆

Eric B. Loucks [a,b,*], Louise Pilote [c], John W. Lynch [d,e], Hugues Richard [c], Nisha D. Almeida [f], Emelia J. Benjamin [g,h,i,1], Joanne M. Murabito [g,h,1]

[a] Department of Community Health, Epidemiology Section, Brown University, Rhode Island, United States
[b] Center for Population Health and Clinical Epidemiology, Brown University, Rhode Island, United States
[c] Department of Medicine, McGill University, Montreal, Quebec, Canada
[d] Sansom Institute, University of South Australia, SA, Australia
[e] Department of Social Medicine, Bristol University, Bristol, United Kingdom
[f] Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Quebec, Canada
[g] Boston University and National Heart Lung and Blood Institute's Framingham Heart Study, Boston, United States
[h] Departments of Medicine and Preventive Medicine, Boston University School of Medicine, Boston, MA, United States
[i] Cardiology Department and Whitaker Cardiovascular Institute, School of Medicine, and Epidemiology, School of Public Health, Boston University, Boston, MA, United States

ARTICLE INFO

Article history:
Available online 23 March 2010

Keywords:
USA
Socioeconomic position
Inflammatory markers
Life course
Social mobility
Heart disease

ABSTRACT

Associations between life course socioeconomic position (SEP) and novel biological risk markers for coronary heart disease such as inflammatory markers are not well understood. Most studies demonstrate inverse associations of life course SEP with C-reactive protein (CRP), interleukin-6 (IL-6) and fibrinogen, however little is known about associations between life course SEP and other inflammatory markers including intercellular adhesion molecule-1 (ICAM-1), tumor necrosis factor II (TNFR2), lipoprotein phospholipase A_2 (Lp-PLA_2) activity, monocyte chemoattractant protein-1 (MCP-1) or P-selectin. The objectives of this analysis were to determine whether three life course SEP frameworks ("accumulation of risk", "social mobility" and "sensitive periods") are associated with the aforementioned inflammatory markers. We examined 1413 Framingham Offspring Study participants (mean age 61.2 ± 8.6 years, 54% women), using multivariable regression analyses. In age- and sex-adjusted regression analyses, cumulative SEP ("accumulation of risk" SEP framework), for low vs. high SEP, was inversely associated with CRP, IL-6, ICAM-1, TNFR2, Lp-PLA_2 activity, MCP-1 and fibrinogen. We found that there were few consistent trends between social mobility trajectories and most inflammatory markers. Own educational attainment was inversely associated with 7 of 8 studied inflammatory markers, while father's education, father's occupation and own occupation were inversely associated with 4, 5 and 4 inflammatory markers, respectively, in age- and sex-adjusted analyses. The strengths of association between SEP and inflammatory markers were typically substantially accounted for by CHD risk markers (smoking, body mass index, systolic blood pressure, total:HDL cholesterol ratio, fasting glucose, medications, depressive symptomatology) suggesting these may be important mechanisms that explain associations between SEP and the studied inflammatory markers.

Introduction

Socioeconomic disparities in coronary heart disease (CHD) exist in many developed countries, where people of lower childhood or adulthood socioeconomic position (SEP) typically have higher risk for incident CHD (Galobardes, Smith, & Lynch, 2006; Gonzalez, Rodriguez Artalejo, & Calero, 1998). There is interest in investigating whether biological risk markers for CHD are related to life course SEP, as a way to evaluate if there is mechanistic evidence consistent with the inverse associations found between SEP and CHD in observational studies. Extensive progress has been made in recent decades on the

☆ The Framingham Heart Study is supported by National Institutes of Health/National Heart, Lung and Blood Institute contract N01-HC-25195. The research was supported by Canadian Institutes of Health operating grant MOP81239 (EBL), National Institutes of Health grants HL064753, HL076784, AG028321 (EJB), and a Canadian Institutes of Health New Investigator Award (EBL). We are grateful to the National Heart, Lung and Blood Institute for providing us with the limited access dataset, in conjunction with Framingham Heart Study investigators who provided other important variables for analyses.

* Corresponding author. Department of Community Health, Epidemiology Section, Brown University, Rhode Island, United States. Tel.: +1 401 863 6283.
E-mail address: eric.loucks@brown.edu (E.B. Loucks).

[1] Authors contributed equally to the study.

0277-9536/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.socscimed.2010.03.012

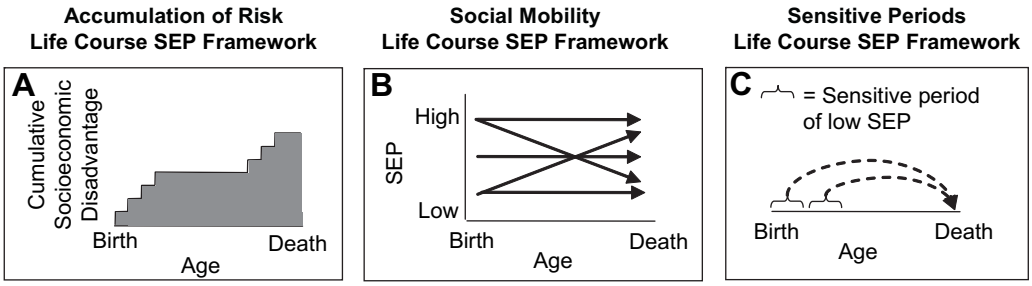


Fig. 1. Conceptual frameworks for three different frameworks to conceptualize socioeconomic position (SEP) across the life course: (A) Accumulation of Risk, (B) Social Mobility, and (C) Sensitive Periods.

involvement of inflammatory processes in atherosclerosis and subsequent risk for CHD, however little is known about associations between life course SEP and many inflammatory markers. Below we introduce major frameworks by which life course SEP is conceptualized, as well as evidence on roles of several inflammatory markers in specific atherosclerotic processes, leading to gaps in knowledge between life course SEP and inflammatory markers.

Galobardes et al. defined several phases for life course SEP including childhood SEP (e.g., parent's education or parent's occupation), young adulthood SEP (e.g., educational attainment), active professional life SEP (e.g., occupation or income), and retirement SEP (e.g., wealth or household conditions) (Galobardes, Shaw, Lawlor, Lynch, & Davey Smith, 2006). A number of frameworks to conceptualize life course SEP have been hypothesized (Fig. 1), including the "accumulation of risk" framework (which focuses on the total cumulative amount of exposure to SEP across the life course), the "social mobility" framework, which recognizes that people have evolving (e.g., increasing, decreasing or stable) SEP across their life span, and the "sensitive periods" framework which suggests that there are certain time periods in the life course when an exposure may have a stronger effect on disease risk than it would during other phases in life (Kuh & Ben-Shlomo, 2004; Kuh, Ben-Shlomo, Lynch, Hallqvist, & Power, 2003).

Novel risk markers for CHD include inflammatory markers. Substantial evidence suggests there is an important inflammatory component in the pathogenesis of atherosclerosis and the pathophysiology is becoming better understood (Libby, 2006). Several inflammatory markers are inversely associated with CHD, including interleukin-6 (IL-6) (Cesari et al., 2003), monocyte chemoattractant protein-1 (MCP-1) (de Lemos et al., 2003), tumor necrosis factor-α (TNF-α) (often measured as soluble tumor necrosis factor receptor II (TNFR2)) (Pai et al., 2004), C-reactive protein (CRP) (Danesh et al., 2004; Ridker et al., 2005), fibrinogen (Keavney et al., 2006), ICAM-1 (Malik et al., 2001), P-selectin (Armstrong, Morrow, & Sabatine, 2006) and lipoprotein-associated phospholipase A_2 (Lp-PLA_2) (Garza et al., 2007). Whether these inflammatory markers are causally related to CHD, rather than simply predictive of CHD, is still under investigation (Elliott et al., 2009; Keavney et al., 2006).

With regard to associations between life course SEP and inflammatory markers, most studies demonstrate inverse associations of life course SEP with CRP, IL-6, fibrinogen and white blood cell count (Brunner et al., 1996; Gimeno et al., 2007, 2008; Koster et al., 2006; Lawlor, Smith, Rumley, Lowe, & Ebrahim, 2005; Loucks et al., 2006; Nazmi & Victora, 2007; Pollitt et al., 2007, 2008; Tabassum et al., 2008), however little is known about associations between life course SEP and other inflammatory markers including TNFR2, MCP-1, ICAM-1, P-selectin and Lp-PLA_2.

The objectives of this study were to determine whether three life course SEP frameworks (i.e., "accumulation of risk", "social mobility", and "sensitive periods" SEP frameworks) are associated with several markers representing diverse inflammatory pathways and processes, including CRP, IL-6, ICAM-1, P-selectin, TNFR2, Lp-PLA_2 activity, MCP-1 and fibrinogen. Further objectives were to evaluate whether CHD risk markers (including smoking, body mass index, systolic blood pressure, total:HDL cholesterol ratio, fasting glucose, cholesterol-lowering medication use, anti-hypertensive medication use and depressive symptomatology) may be explanatory mechanisms for any observed associations between life course SEP and inflammatory markers. It should be emphasized, as discussed by Hallqvist et al. that it is likely not possible to critically test individual contributions of accumulation of risk *vs.* social mobility *vs.* sensitive periods due to mutual confounding between the three SEP frameworks (Hallqvist, Lynch, Bartley, Lang, & Blane, 2004). Consequently, we do not propose to statistically compare the contributions of each of these SEP frameworks to each other. As triangulation of methodological approaches enables a more thorough understanding of health determinants, we utilized the three life course SEP framework to offer three approaches to evaluate the potential association of life course SEP with inflammatory markers. Utilizing information from all three SEP frameworks will provide a more complete picture of life course SEP determinants of inflammatory markers than if findings were presented on only one of the life course SEP frameworks.

Materials and methods

Study sample

The Framingham Heart Study is a community-based observational cohort that was initiated in 1948. The Framingham Offspring Study began in 1971 with recruitment of 5124 men and women who were offspring (or offspring's spouses) of the Original Framingham Heart Study participants. The design and selection criteria of the Framingham Offspring Study have been described elsewhere (Kannel, Feinleib, McNamara, Garrison, & Castelli, 1979). All study participants received routine medical history and physical examinations, laboratory assessments of cardiovascular risk factors, and anthropometric measurements approximately every 3–4 years. Framingham participants signed informed consent and the Framingham Study is reviewed annually by the Boston University Medical Center Institutional Review Board.

Father's education was measured directly from fathers in the Original Cohort of the Framingham Heart Study. All other variables were measured in the Offspring Cohort. Own occupation was measured in Offspring Examination 2 (1979–1982) and Examination 7 (1998–2001). Own education was measured in Examination 2 and Examination 3 (1984–1987). Inflammatory markers and covariates were measured during Examination 7. There were 3475 Framingham Offspring Study participants in the National Heart, Lung and Blood Institute (NHLBI) data repository dataset who completed Examination 7. Of these, 1451 participants did not have fathers in the Original Framingham Study (362 participants had

only a mother in the Original Cohort, and 1089 participants were spouses of participants in the Offspring Cohort and had no father in the Original Cohort) from which measures of father's education were directly obtained. In order to be eligible for the Framingham Offspring study, participants needed to be offspring of a male or female Original Framingham Study participant, or a spouse of the offspring. An additional 78 participants had fathers who were missing the education variable. A further 180 participants were missing own education or occupation variables. We further restricted the participants to those ≥28 years when education and occupation were measured, resulting in 97 participants being excluded. Restricting participants to ≥28 years was done in order to allow at least 10 years from expected graduation of high school to obtain further education and become established in an occupational category. An additional 94 subjects with prevalent CVD were excluded, resulting in 1575 participants. Of these, the number of participants missing inflammatory marker variables were as follows: CRP: 62; P-selectin: 63; ICAM-1: 66; MCP-1: 99; interleukin-6: 65; TNFR2: 103; and Lp-PLA_2 activity: 64. As CRP is the most widely recognized inflammatory marker to date with regard to being a potential risk marker for CHD, analyses report sample sizes for CRP analyses ($n = 1513$).

Analyses on excluded versus included participants found that excluded and included participants were similar with regard to smoking, body mass index, fasting glucose, IL-6, TNFR, Lp-PLA_2 activity, P-selectin, fibrinogen, cholesterol-lowering medication use, and depressive symptomatology. Excluded participants were more likely to be older ($p < 0.001$), have higher systolic blood pressure ($p = 0.001$), CRP ($p = 0.008$), ICAM-1 ($p = 0.001$), MCP-1 ($p = 0.005$), and lower total:HDL cholesterol ratio ($p = 0.04$), compared with included participants. Furthermore, excluded participants were more likely than included participants to be taking anti-hypertensive medication ($p = 0.006$), have ≤high school education ($p < 0.0001$), have a manual occupation ($p = 0.001$), and have fathers with ≥high school education ($p = 0.009$).

Independent variables

Childhood SEP (father's education and father's occupation)

Childhood SEP was measured in primary analyses by father's educational attainment, obtained directly from Offspring cohort participants' fathers who were enrolled in the Original Framingham Heart Study. Father's educational level was measured at enrollment between 1948 and 1950 when their mean age was 44 years (range: 28–62), categorized into 3 groups: <high school ($n = 799$), high school ($n = 418$) and >high school ($n = 452$). Sensitivity analyses used father's occupation measured during Original Framingham Heart Study Examination 6 where all participants were greater than age 28 years. Father's occupation was categorized identically to Offspring occupation described below.

Young adulthood SEP (own educational attainment)

Own education was measured directly from Framingham Offspring Study participants at Examination 3 (1984–1987); if Examination 3 education was missing, then Examination 2 assessment (1979–1982) was used ($n = 197$). Education was categorized into 3 groups: ≤12 ($n = 612$), 13–16 ($n = 719$) and ≥17 ($n = 338$) years education. Pearson's correlation coefficient for associations between educational attainment at Examination 2 and Examination 3 was 0.89.

Active professional life SEP (own occupation)

Own occupation was measured from Offspring participants at Examination 7 (age range: 41–81 years). If participants were coded as missing, retired, unemployed, or housewife in Examination 7, then occupation from Examination 2 (1979–1982) was used ($n = 685$; age range: 28–67 years; mean age: 47.7 years). Those who were missing occupation at Examinations 2 and 7 were excluded from analyses. No participants were coded as retired or unemployed at both exams 2 and 7. Thus, 984 had occupation data from Examination 7, and the remaining 685 participants utilized occupation data from Examination 2. Included participants were age ≥28 years at the time of occupation assessment in order to allow at least 10 years between typical age of high school graduation to complete education and obtain an occupation. Occupation was categorized as laborer ($n = 306$), clerical/sales ($n = 203$), housewife ($n = 375$), technical/supervisor ($n = 195$), and executive/professional ($n = 590$). These occupation categorizations were broadened to only two or three categories for analyses, as detailed in the SEP Frameworks section below. Some misclassification of occupation is expected due to the wide age range for when it was measured, as described in more detail in the Discussion section.

SEP frameworks

Analyses that used an **accumulation of risk** framework used a cumulative SEP score (range: 0–6) including father's education (<high school = 0, high school = 1, >high school = 2), own education (≤12 years = 0, 13–16 years = 1, ≥17 years = 2) and own occupation (laborer = 0, clerical/sales/housewife = 1, executive/professional/supervisory/technical = 2). Higher cutpoints were used for own educational categories, compared with father's (i.e., own education categorized as: ≤12, 13–16 and ≥17 years education; father's education categorized as: <high school, high school and >high school) to account for secular trends of increased normative levels of education across generations. Analyses that used a **social mobility** framework utilized dichotomous categories of father's education (lower: <high school, higher: ≥high school) and own occupation (lower: laborer, higher: housewife/clerical/sales/supervisory/technical/professional/executive). Analyses tested the association of four possible types of social mobility across the life course: stable high SEP (high childhood and adulthood SEP), decreasing SEP (high childhood SEP and low adulthood SEP), increasing SEP (low childhood SEP and high adulthood SEP), and stable low SEP (low childhood and adulthood SEP). Using the occupation categorizations described above, we expect there will be SEP misclassification, however given the large variety of occupations in society, and the contributing factors to occupation-based SEP (such as income and social prestige), this appeared to be an acceptable approximation of occupation-based SEP that took into account factors such as income, social prestige and educational requirements. Analyses assessing the **sensitive periods** framework assessed associations between each individual SEP measure (i.e., father's education, own education, own occupation) and inflammatory marker concentrations, further adjusting for all SEP measures other than the independent SEP variable.

Dependent variables

Inflammatory markers were measured during Examination 7 (1998–2001). Fasting morning serum samples were collected then stored at −80 °C. Serum CRP was measured once using a high-sensitivity assay (Dade Behring BN100 nephelometer, Deerfield, IL; inter-assay CV 3.2%). IL-6, ICAM-1, TNFR2, MCP-1, and P-selectin concentrations were measured in duplicate and averaged using commercially-available Enzyme-Linked Immunoassay kits (R&D Systems, Minneapolis, MN) following previously described quality control procedures (Keaney et al., 2003, 2004). Lp-PLA_2 activity was measured by DiaDexus, Inc., San Francisco, CA. Biomarker measurement reproducibility was good (intra-assay coefficients of

Table 1
Age and sex-adjusted characteristics (95% confidence intervals for the mean) according to cumulative socioeconomic position (SEP) score.

	Cumulative SEP score		
	0–1 ($n = 343$)	2–3 ($n = 553$)	4–6 ($n = 617$)
Clinical characteristics			
Age, years	63.7 (62.8,64.6)	58.8 (58.1,59.4)	57.9 (57.3,58.6)
Female, %	56.6 (51.1,61.9)	59.9 (55.6,64.0)	46.0 (42.0,50.1)
Current smoker, %	20.1 (16.1,24.9)	14.0 (11.4,17.1)	8.4 (6.5,10.8)
Body mass index, kg/m^2	29.0 (28.4,29.5)	28.5 (28.1,28.9)	27.7 (27.3,28.1)
Systolic blood pressure, mmHg	128 (126,129)	125 (124,126)	125 (124,126)
Total:HDL[a] cholesterol ratio	4.1 (4.0,4.3)	4.2 (4.1,4.3)	4.0 (3.9,4.1)
Fasting glucose, mg/dL	105 (102,108)	104 (102,106)	102 (100,104)
Taking cholesterol-lowering medication, %	23.9 (19.7,28.6)	16.4 (13.5,19.7)	15.7 (131,18.8)
Taking anti-hypertensive medication, %	32.8 (28.0,38.0)	31.3 (27.6,35.4)	27.1 (23.7,30.9)
Depressive symptoms, CES-D[a] score	5.7 (5.1,6.4)	5.8 (5.3,6.3)	4.6 (4.1,5.1)
Prevalent cardiovascular disease, %	5.0 (3.2,7.6)	3.5 (2.3,5.3)	2.3 (1.5,3.7)
Inflammatory markers			
CRP, mg/L[a]	2.57 (2.29,2.88)	2.26 (2.06,2.47)	1.90 (1.74,2.07)
IL-6, pg/mL[a]	3.10 (2.89,2.33)	2.88 (2.72,3.04)	2.69 (2.55,2.83)
ICAM-1, ng/mL[a]	252 (246,259)	249 (244,253)	236 (233,240)
TNFR2, pg/mL[a]	2074 (2013,2136)	2064 (2016,2112)	1937 (1894, 1980)
Lp-PLA$_2$ activity, nmol/min/mL[a]	140 (137,143)	140 (138,143)	136 (133,138)
P-Selectin, ng/mL[a]	34.7 (33.4,36.0)	36.5 (35.4,37.6)	34.8 (33.8,35.8)
MCP-1, pg/mL[a]	316 (305,327)	304 (296,313)	299 (291,307)
Fibrinogen, g/L[a]	382 (374,389)	371 (365,376)	367 (362,372)

[a] Mean inflammatory concentrations were obtained using linear regression on log-transformed variables, and then back-transformed to obtain estimates of the mean concentrations.

variation for IL-6, 3.1%; ICAM-1, 3.7%; TNFR2, 2.2%; MCP-1, 3.8%; P-selectin, 3.0%, Lp-PLA$_2$ activity, 4.3%).

Covariates

CHD risk markers were measured at Examination 7 (1998–2001). Current cigarette smoking was determined by self-report if it occurred regularly in the past year. Systolic blood pressure was calculated as the average of the clinic physician's two seated systolic blood pressure measurements. Body mass index was calculated as the weight in kilograms divided by the square of the height in meters (kg/m^2). Fasting glucose was measured with a hexokinase reagent kit (A-gent glucose test, Abbott, South Pasadena, California); intra-assay coefficients of variation ranged from 2% to 3%. High density lipoprotein and total cholesterol concentrations were measured by automated enzymatic techniques (McNamara & Schaefer, 1987). Depressive symptomatology was measured using the Center for Epidemiologic Studies Depression (CES-D) scale, and was analyzed as a continuous variable. Medication use was classified by self-report.

Statistical analysis

Sex- and age-standardized descriptive statistics (means and proportions) were generated for CHD risk factors (systolic blood pressure, total:HDL cholesterol ratio, fasting glucose, body mass index, smoking and anti-hypertensive medication use) and prevalent CVD, according to cumulative SEP score.

Due to non-normality in the distributions, the inflammatory markers were natural log-transformed. Age- and sex-adjusted mean inflammatory marker concentrations were obtained using linear regression on natural log-transformed variables, then back-transformed to obtain estimates of the mean concentrations.

Multivariable regression analyses evaluated associations of life course SEP with log-transformed inflammatory marker concentrations. Analyses adjusted for potential confounders including age and sex, as well as CHD risk markers including smoking, body mass index, systolic blood pressure, total:HDL cholesterol ratio, fasting glucose, cholesterol-lowering medication use, anti-hypertensive medication use and depressive symptomatology. The three SEP variables (father's education, own education and own occupation) were not correlated highly enough to be of concern to simultaneously adjust for all three in a single multivariable model (correlation coefficients ranged from 0.27 to 0.51). Consequently, all three measures of SEP were simultaneously adjusted for in analyses testing the sensitive periods framework. Generalized Estimating Equations (GEE) were used to account for clustering of outcomes by family. There were 1051 clusters, with the minimum cluster size of 1 (indicating these participants are not part of a cluster) and maximum cluster size of 6. There was no evidence of effect modification by sex, consequently sexes were pooled. Analyses were conducted using the statistical program SAS version 9.1(SAS Institute, Cary, NC).

Results

Fifty-four percent of participants were women, and the mean age was 61.2 (8.6 SD) years. Higher cumulative SEP levels were associated with more favorable cardiovascular risk factors including smoking, body mass index, systolic blood pressure, fasting glucose, cholesterol-lowering medications, anti-hypertensive medications, and depressive symptomatology in age- and sex-adjusted analyses (Table 1). There were no associations between cumulative SEP score and total:HDL cholesterol ratio. Those with a low cumulative SEP score (0 or 1) were more likely to be older and more likely to be female (women tended to have lower educational attainment than men, and fewer females were in very high level occupations; data not shown) than participants with a high SEP score (4–6) (Table 1). Age- and sex-adjusted analyses describing the logarithmically back-transformed mean inflammatory marker concentrations showed inverse associations of cumulative SEP with CRP, IL-6, ICAM-1, TNFR2, Lp-PLA$_2$ activity, MCP-1 and fibrinogen (Table 1). There was an inverse U association between cumulative SEP and P-selectin. Effect sizes varied depending on inflammatory marker, shown in detail in Table 1. For example, participants with high cumulative SEP scores (SEP score = 4–6) had lower mean CRP concentrations (1.99, 95% CI: 1.81, 2.19 mg/L) than participants with low cumulative SEP scores

Table 2
Multivariable regression analyses demonstrating associations of cumulative life course socioeconomic position (SEP) score with inflammatory marker concentrations.

Marker	Cumulative SEP score	N	Model adjustment	
			Age, sex	Age, sex, CHD risk markers[a]
			β (95% CI)	β (95% CI)
CRP, ln mg/L	0–1	343	**0.289 (0.140,0.439)**	0.117 (−0.021,0.254)
	2–3	553	**0.179 (0.051,0.307)**	0.064 (−0.050,0.177)
	4–6	617	0.00	0.00
IL-6, ln pg/mL	0–1	342	**0.135 (0.044,0.227)**	0.042 (−0.047,0.131)
	2–3	553	0.072 (−0.011,0.156)	0.013 (−0.067,0.092)
	4–6	615	0.00	0.00
ICAM-1, lnng/mL	0–1	343	**0.067 (0.033,0.100)**	0.030 (−0.002,0.062)
	2–3	552	**0.047 (0.020,0.075)**	0.025 (−0.000,0.051)
	4–6	614	0.00	0.00
TNFR2, ln pg/mL	0–1	338	**0.070 (0.033,0.107)**	**0.050 (0.013,0.086)**
	2–3	538	**0.062 (0.030,0.093)**	**0.046 (0.015,0.077)**
	4–6	596	0.00	0.00
Lp-PLA_2, lnnmol/min/mL	0–1	342	**0.037 (0.006,0.068)**	0.018 (−0.010,0.045)
	2–3	553	**0.032 (0.006,0.058)**	0.012 (−0.012,0.037)
	4–6	616	0.00	0.00
P-Selectin, lnng/mL	0–1	343	0.004 (−0.046,0.055)	−0.031 (−0.081,0.019)
	2–3	553	**0.048 (0.007,0.090)**	0.029 (−0.012,0.070)
	4–6	616	0.00	0.00
MCP-1, ln pg/mL	0–1	338	**0.051 (0.006,0.096)**	0.035 (−0.011,0.081)
	2–3	542	0.020 (−0.020,0.061)	0.010 (−0.031,0.051)
	4–6	609	0.00	0.00
Fibrinogen, ln g/L	0–1	344	**0.035 (0.010,0.060)**	0.015 (−0.009,0.038)
	2–3	553	0.013 (−0.008,0.035)	−0.005 (−0.025,0.016)
	4–6	614	0.00	0.00

Bold values denote where the 95% confidence intervals do not encompass the reference category.
[a] CHD risk markers include smoking, body mass index, systolic blood pressure, total:HDL cholesterol ratio, fasting glucose, cholesterol-lowering medication use, antihypertensive medication use and depressive symptomatology.

(SEP score = 0 or 1; mean CRP concentration = 2.47, 95% CI: 2.22, 2.75 mg/L).

In regression analyses testing the **accumulation of risk SEP framework**, cumulative SEP was inversely associated with CRP, IL-6, ICAM-1, TNFR2, Lp-PLA_2 activity, MCP-1 and fibrinogen after adjusting for age and sex (Table 2). There was an inverse U association between cumulative SEP and P-selectin. Further adjustment for CHD risk markers substantially reduced effect sizes.

In analyses testing the **social mobility SEP framework**, participants with declining SEP across their life course (as measured by father's education and own occupation) had elevated concentrations of IL-6, ICAM-1, and TNFR2, compared with participants who had high SEP in childhood and adulthood after adjusting for age and sex (Table 3). Furthermore, participants who had low SEP in both childhood and adulthood were more likely to have elevated CRP and TNFR2 compared with those who had high SEP in childhood and adulthood in age- and sex-adjusted analyses. Participants who experienced low childhood SEP and high adulthood SEP were more likely to have elevated ICAM-1, TNFR2, and Lp-PLA_2 activity compared with participants who had high SEP in childhood and adulthood, after adjusting for age and sex (Table 3). Further adjustment for CHD risk markers typically accounted for a substantial amount of the strength of the associations of SEP with inflammatory markers.

For multivariable regression analyses evaluating the **sensitive periods SEP framework**, there were inverse associations of own education with 7 of the 8 investigated inflammatory markers (CRP, IL-6, ICAM-1, TNFR2, P-selectin, MCP-1 and fibrinogen) after adjusting for age and sex (Table 4). Father's education an down occupation were inversely associated with 4 of the 8 explored inflammatory markers (father's education inversely associated with CRP, ICAM-1, TNFR2 and Lp-PLA_2 activity; own occupation inversely associated with CRP, IL-6, ICAM-1 and TNFR2). Adjustment for all SEP measures demonstrated that educational attainment had evidence of inverse associations with several inflammatory markers (CRP, ICAM-1, P-selectin, MCP-1 and fibrinogen) independently of other SEP measures (Table 4). There was little evidence of associations of father's education or own occupation with most inflammatory markers independent of other SEP measures. Further adjustment for CHD risk markers typically accounted for a substantial amount of the strength of association between SEP measures and inflammatory markers (Table 4).

Sensitivity analyses evaluated associations of the life course SEP frameworks with inflammatory markers, where father's occupation was used as a measure of childhood SEP instead of father's education. Findings were generally similar, with point estimates typically somewhat stronger when using father's occupation (Tables S1–S3). Father's occupation was inversely associated with CRP, IL-6, ICAM-1, TNFR2 and MCP-1 in analyses adjusted for age and sex. Associations remained for father's occupation with CRP, IL-6 and TNFR2 after further adjustment for own education and occupation (Table S3). Additional adjustment for CHD risk markers typically accounted for a substantial amount of the association strength between father's occupation and inflammatory markers (Table S3).

Discussion

This study provided evidence that cumulative life course SEP is inversely associated with many inflammatory markers including CRP, IL-6, ICAM-1, TNFR2, LP-PLA_2 activity, MCP-1 and fibrinogen in age- and sex-adjusted analyses. Own education was associated with almost all studied inflammatory markers (CRP, IL-6, ICAM-1, TNFR2, P-selectin, MCP-1 and fibrinogen), while father's education, father's occupation and own occupation were associated with several but not all inflammatory markers in age- and sex-adjusted

Table 3
Multivariable regression analyses demonstrating the association of the *social mobility* framework of life course socioeconomic position (SEP) with inflammatory marker concentrations.

Marker	SEP level in childhood/adulthood	N	Model adjustment	
			Age, sex	Age, sex, CHD risk markers[a]
			β (95% CI)	β (95% CI)
CRP, ln mg/L	Low/Low	167	**0.241 (0.050,0.432)**	0.116 (−0.067,0.299)
	High/Low	90	0.246 (−0.008,0.499)	0.162 (−0.046,0.369)
	Low/High	538	0.146 (0.018,0.275)	0.005 (−0.109,0.120)
	High/High	718	0.00	0.00
Interleukin-6, ln pg/mL	Low/Low	166	0.067 (−0.050,0.183)	−0.002 (−0.114,0.111)
	High/Low	90	**0.214 (0.052,0.376)**	0.123 (−0.021,0.267)
	Low/High	538	0.047 (−0.035,0.129)	−0.016 (−0.094,0.062)
	High/High	716	0.00	0.00
ICAM-1, lnng/mL	Low/Low	167	0.039 (−0.002,0.080)	0.013 (−0.027,0.053)
	High/Low	90	**0.097 (0.040,0.154)**	**0.057 (0.008,0.106)**
	Low/High	537	**0.031 (0.001,0.061)**	0.010 (−0.018,0.038)
	High/High	715	0.00	0.00
TNFR2, ln pg/mL	Low/Low	161	**0.087 (0.039,0.135)**	**0.067 (0.020,0.115)**
	High/Low	87	**0.102 (0.046,0.159)**	**0.087 (0.029,0.145)**
	Low/High	532	**0.035 (0.001,0.069)**	0.009 (−0.024,0.042)
	High/High	692	0.00	0.00
Lp-PLA_2, activity lnnmol/min/mL	Low/Low	166	0.035 (−0.006,0.075)	0.026 (−0.011,0.062)
	High/Low	90	0.015 (−0.035,0.065)	0.019 (−0.029,0.067)
	Low/High	538	**0.030 (0.001,0.059)**	0.002 (−0.025,0.028)
	High/High	717	0.00	0.00
P-Selectin, lnng/mL	Low/Low	166	−0.010 (−0.073,0.053)	−0.042 (−0.102,0.019)
	High/Low	90	0.025 (−0.041,0.090)	0.007 (−0.056,0.070)
	Low/High	539	−0.021 (−0.065,0.023)	−0.042 (−0.085,0.001)
	High/High	717	0.00	0.00
MCP-1, ln pg/mL	Low/Low	165	0.015 (−0.046,0.077)	0.001 (−0.062,0.064)
	High/Low	87	0.051 (−0.019,0.121)	0.033 (−0.039,0.106)
	Low/High	531	0.002 (−0.040,0.044)	−0.006 (−0.048,0.037)
	High/High	706	0.00	0.00
Fibrinogen, ln g/L	Low/Low	167	0.016 (−0.016,0.048)	0.004 (−0.026,0.034)
	High/Low	90	0.034 (−0.006,0.074)	0.022 (−0.014,0.057)
	Low/High	539	0.012 (−0.010,0.034)	−0.009 (−0.030,0.012)
	High/High	715	0.00	0.00

Bold values denote where the 95% confidence intervals do not encompass the reference category.
[a] CHD risk markers include smoking, body mass index, systolic blood pressure, total:HDL cholesterol ratio, fasting glucose, cholesterol-lowering medication use, anti-hypertensive medication use and depressive symptomatology.

analyses (father's education: CRP, ICAM-1, TNFR2, Lp-PLA_2 activity; father's occupation: CRP, IL-6, ICAM-1, TNFR2 and MCP-1; offspring occupation: CRP, IL-6, ICAM-1 and TNFR2). There were minimal consistent trends between social mobility trajectories and the most inflammatory markers, with evidence that low SEP at any time in the life course conferred risk for elevated inflammatory markers. The associations between the SEP measures and inflammatory markers were typically substantially accounted for by adjusting for CHD risk markers, suggesting these may be important mechanisms that explain some of the association between SEP and the studied inflammatory markers.

Prior literature

The association of adulthood SEP with inflammatory markers has been well studied for CRP and fibrinogen, and less so for other inflammatory markers addressed in this report. A systematic review showed that 20 of 21 studies demonstrated inverse associations between adulthood SEP and CRP in minimally adjusted analyses. Further adjustment for demographic, anthropometric and other covariates typically reduced strengths of association, and only 9 of 16 multivariate-adjusted studies showed significant inverse associations (Nazmi & Victora, 2007). Inverse associations of adulthood SEP with IL-6 and fibrinogen have been found in most studies using large sample sizes ($n > 1000$), with effect sizes typically substantially reduced after adjusting for variables that may be partly on the causal pathway, such as smoking and obesity (Brunner et al., 1996; Gimeno et al., 2007; Koster et al., 2006; Loucks et al., 2006; Ramsay et al., 2008). These findings are in general agreement with those reported in the present study. Little is known about the association between adulthood SEP and other inflammatory markers addressed in this article including TNFR2, MCP-1, ICAM-1, P-selectin and Lp-PLA_2. For associations of SEP with tumor necrosis factor, in two large studies ($n > 1000$), adulthood SEP was shown to be inversely associated with TNF-α concentrations in multivariable-adjusted analyses (Koster et al., 2006; Panagiotakos et al., 2005). We showed previously in the Framingham Offspring study that educational attainment was inversely associated with ICAM-1 and MCP-1 in multivariable-adjusted analyses (Loucks et al., 2006). Analyses reported in our current paper were in general support of other studies, which showed inverse associations of education and occupation with CRP, IL-6, ICAM-1, and TNFR2 concentrations after adjusting for age and sex, and general reductions in effect size after further adjusting for CHD risk markers.

With regard to associations between childhood SEP and inflammatory marker concentrations, less is known on this topic compared with adulthood SEP and inflammatory markers. However, a reasonable number of studies showed inverse

Table 4
Multivariable regression analyses for the association of the *sensitive periods* framework of life course socioeconomic position with concentrations of inflammatory markers.

	Model adjustment			
	Age, sex	Age, sex, other SEP measures[b]	Age, sex, CHD risk markers[c]	Age, sex, other SEP measures,[b] CHD risk markers[c]
Inflammatory marker	β[a]	β	β	β
Father's education				
CRP, ln mg/L	**0.163 (0.023,0302)**	0.098 (−0.048,0.244)	0.017 (−0.108,0.142)	−0.008 (−0.138,0.121)
IL-6, ln pg/mL	0.067 (−0.021,0.155)	0.031 (−0.060,0.123)	−0.002 (−0.086,0.082)	−0.015 (−0.102,0.072)
ICAM-1, lnng/mL	**0.038 (0.006,0.070)**	0.020 (−0.014,0.055)	0.009 (−0.021,0.039)	0.002 (−0.030,0.033)
TNFR2, ln pg/mL	**0.055 (0.018,0.093)**	**0.044 (0.004,0.084)**	0.028 (−0.009,0.064)	0.019 (−0.020,0.058)
Lp-PLA_2, lnnmol/min/mL	**0.036 (0.005,0.068)**	0.030 (−0.003,0.063)	0.004 (−0.025,0.032)	0.000 (−0.030,0.030)
P-Selectin, lnng/mL	0.003 (−0.044,0.050)	−0.014 (−0.063,0.036)	−0.027 (−0.074,0.020)	−0.033 (−0.082,0.015)
MCP-1, ln pg/mL	0.016 (−0.029,0.061)	−0.009 (−0.058,0.039)	−0.003 (−0.043,0.049)	−0.015 (−0.064,0.034)
Fibrinogen, ln g/L	0.017 (−0.007,0.041)	0.010 (−0.014,0.035)	−0.005 (−0.027,0.017)	−0.008 (−0.031,0.015)
Own education				
CRP, ln mg/L	**0.336 (0.179, 0.494)**	**0.249 (0.072,0.426)**	**0.149 (0.010,0.288)**	0.111 (−0.063,0.286)
IL-6, ln pg/mL	**0.148 (0.051, 0.245)**	0.102 (−0.010,0.214)	0.049 (−0.045,0.142)	0.028 (−0.081,0.137)
ICAM-1, lnng/mL	**0.064 (0.030, 0.098)**	**0.040 (0.002,0.079)**	0.027 (−0.005,0.058)	0.013 (−0.023,0.049)
TNFR2, ln pg/mL	**0.048 (0.009, 0.088)**	0.012 (−0.032,0.056)	0.025 (−0.014,0.064)	−0.004 (−0.048,0.040)
Lp-PLA_2, lnnmol/min/mL	0.024 (−0.008, 0.055)	0.010 (−0.027,0.046)	−0.000 (−0.030,0.029)	−0.013 (−0.048,0.022)
P-Selectin, lnng/mL	**0.073 (0.023, 0.123)**	**0.071 (0.014,0.128)**	0.034 (−0.016,0.084)	0.037 (−0.019,0.093)
MCP-1, ln pg/mL	**0.062 (0.013, 0.111)**	**0.071 (0.018,0.124)**	0.039 (−0.012,0.089)	0.052 (−0.002,0.106)
Fibrinogen, ln g/L	**0.037 (0.012, 0.062)**	**0.031 (0.003,0.060)**	0.015 (−0.009,0.039)	0.014 (−0.013,0.042)
Own occupation				
CRP, ln mg/L	**0.221 (0.066, 0.375)**	0.123 (−0.049,0.296)	**0.167 (0.028,0.307)**	0.150 (−0.004,0.304)
IL-6, ln pg/mL	**0.124 (0.026, 0.223)**	0.087 (−0.024,0.198)	0.065 (−0.027,0.158)	0.065 (−0.040,0.170)
ICAM-1, lnng/mL	**0.057 (0.022, 0.093)**	0.039 (−0.001,0.079)	0.032 (−0.001,0.066)	0.029 (−0.008,0.066)
TNFR2, ln pg/mL	**0.076 (0.037, 0.114)**	**0.064 (0.023,0.106)**	**0.070 (0.031,0.109)**	**0.069 (0.026,0.111)**
Lp-PLA_2, lnnmol/min/mL	0.019 (−0.013, 0.050)	0.008 (−0.027,0.0432)	0.027 (−0.003,0.057)	0.030 (−0.004,0.064)
P-Selectin, lnng/mL	0.029 (−0.020, 0.077)	0.012 (−0.040,0.064)	0.017 (−0.030,0.063)	0.017 (−0.033,0.067)
MCP-1, ln pg/mL	0.024 (−0.025, 0.073)	−0.006 (−0.056,0.045)	0.007 (−0.045,0.058)	0.015 (−0.067,0.038)
Fibrinogen, ln g/L	0.020 (−0.006, 0.046)	0.009 (−0.020,0.038)	0.013 (−0.011,0.038)	0.012 (−0.016,0.039)

Bold values denote where the 95% confidence intervals do not encompass the reference category.

[a] For father's education, regression coefficients (β) represent the relative logarithmic change in biomarker units (95% confidence intervals) for participants with father's education of <high school, versus father's education of >high school. Multivariable regressions included category of high school education (results not shown). For own education, regression coefficients (β) represent the relative logarithmic change in biomarker units (95% confidence intervals) for participants with own education of ≤12 years, versus own education of >16 years. For own occupation, regression coefficients (β) represent the relative logarithmic change in biomarker units (95% confidence intervals) for laborers versus participants with supervisory/technical/professional/executive jobs.

[b] Other SEP measures correspond to the two SEP variables other than the independent variable. For example, if father's education is the independent variable, other SEP measures would include own education and occupation.

[c] CHD risk markers include smoking, body mass index, systolic blood pressure, total:HDL cholesterol ratio, fasting glucose, cholesterol-lowering medication use, anti-hypertensive medication use and depressive symptomatology.

associations of childhood SEP (typically measured as father's occupation or father's education) with CRP concentrations (e.g., (Lawlor et al., 2005; Tabassum et al., 2008)) and at least 2 studies demonstrated no association (Gimeno et al., 2008; Pollitt et al., 2007). Our findings supported others' demonstrated inverse associations between childhood SEP and CRP as well as fibrinogen in age- and sex-adjusted analyses (Brunner et al., 1996; Pollitt et al., 2007; Tabassum et al., 2008). Our findings demonstrated father's education and father's occupation were inversely associated with a trend in fibrinogen concentrations, but the 95% confidence intervals for the effect point estimates just encompassed the referent point estimate, suggesting a null or weak association. With regard to other inflammatory markers addressed in this report (i.e., IL-6, ICAM-1, TNFR2, Lp-PLA_2, P-selectin and MCP-1), very little is known about their associations with childhood SEP. Mendall et al. showed that father's occupational prestige was inversely associated with mean IL-6 concentrations in 198 British men aged 50–69 years after adjustment for several CHD risk factors; no association was found with TNF-α (Mendall et al., 1997). To our knowledge, very little is known about association of childhood SEP with ICAM-1, TNFR2, P-selectin, MCP-1 and Lp-PLA_2 activity. We found inverse associations of father's education with ICAM-1, TNFR2 and Lp-PLA_2 activity; strengths of association were reduced after adjusting for adulthood SEP and CHD risk markers, suggesting these may be mechanisms accounting for the association.

Cumulative SEP studies to date typically showed inverse associations between cumulative life course SEP with CRP and fibrinogen concentrations (Gimeno et al., 2008; Lawlor et al., 2005; Pollitt et al., 2008; Tabassum et al., 2008). Our findings are consistent with other observed inverse associations of cumulative SEP with CRP and fibrinogen concentrations. Little prior information is available on associations between cumulative life course SEP and concentrations of other inflammatory markers addressed in this report (i.e., IL-6, ICAM-1, TNFR2, Lp-PLA_2, P-selectin and MCP-1).

Furthermore, to our knowledge, there is very minimal information on associations between social mobility and most inflammatory marker concentrations reported here. In conjunction with findings from "accumulation of risk" and "sensitive periods" SEP frameworks discussed above, this paper's analyses found general evidence that low SEP at any time in the life course was associated with elevated inflammatory markers. For example, in social mobility analyses there were associations between declining life course SEP and elevated IL-6, ICAM-1 and TNFR2 in age- and sex-adjusted analyses. Further, participants with low childhood SEP and high adulthood SEP had elevated ICAM-1, TNFR2 and Lp-PLA_2 activity; participants with low childhood SEP and low adulthood SEP had elevated CRP and TNFR2. Findings were not as consistent for social mobility analyses as for the other two SEP frameworks, likely in part because only father's education and offspring occupation were used in these analyses, rather than all three sensitive

periods (including own education which was strongly associated with most inflammatory markers).

Mechanisms

For the observed associations between SEP and inflammatory markers in this study after adjusting for age and sex, further adjustment for CHD risk markers typically accounted for a substantial amount of the association. These factors may be substantial mediating pathways through which life course SEP influences inflammatory markers. There is evidence that SEP in childhood and adulthood may influence health behaviors and CHD risk markers, which could serve as intermediate mechanisms contributing to the emergence of altered concentrations of inflammatory markers. For example, childhood SEP is inversely associated with obesity in adult females (Senese, Almeida, Kittler Fath, Smith, & Loucks, 2009) and smoking in adult males and females (Lawlor, Batty, et al., 2005). With regard to adulthood SEP, those with low SEP tend to have higher smoking rates (Gilman et al., 2008), blood pressure (Colhoun, Hemingway, & Poulter, 1998), diabetes (Maty, Everson-Rose, Haan, Raghunathan, & Kaplan, 2005), and in the case of women, obesity (McLaren, 2007). Cigarette smoking and obesity have been shown to be associated with inflammatory markers reported in this cohort and may be mechanisms by which childhood SEP influences inflammatory marker concentrations. An additional potential mechanism is through psychological distress, such as depression or anxiety. Childhood SEP is associated with depression in youth and adulthood (Gilman, Abrams, & Buka, 2003). Adulthood SEP is also related to depression (Muntaner, Eaton, Miech, & O'Campo, 2004). Reports have demonstrated that depression is associated with elevated inflammatory markers including C-reactive protein, IL-6, and ICAM-1 (Empana et al., 2005). The causal direction between depression and inflammation is still being elucidated (Almeida et al., 2009; Dantzer, O'Connor, Freund, Johnson, & Kelley, 2008). Furthermore, low SEP is associated with chronic inflammatory conditions such as periodontal disease that can influence circulating levels of inflammatory markers including CRP (Borrell & Crawford, 2008; Linden, McClean, Young, Evans, & Kee, 2008).

Strengths and weaknesses

Strengths of the present investigation include the well-characterized community-based cohort in which a wide range of CHD risk markers were routinely measured using high quality methods, enabling us to adjust for a large number of covariates. Furthermore, childhood SEP was obtained directly from parents, which can limit recall bias compared to studies that asked participants to estimate parents' SEP.

With regard to weaknesses, because the historical design of the Framingham Offspring Study reflected the population of Framingham, Massachusetts at study onset in 1948, the Original and Offspring cohorts are largely composed of participants of European descent. Consequently, the generalizability of our findings to other races/ethnicities is uncertain. Additionally, excluded participants were more likely to be of low SEP, and have elevated CHD risk markers, certain inflammatory markers and CHD, compared with included participants. We expect this could have induced bias in the reported findings likely towards the null as the excluded participants tended to have extreme values for both the exposures and outcomes. Furthermore, our primary analyses on life course SEP included only 3 measures: father's education, own education and own occupation. Other studies that have additional measures of life course SEP (such as *in utero* SEP, or multiple measures of occupation throughout the life course) will provide richer data as to the potential contribution of sensitive periods, accumulation of risk and social mobility SEP frameworks to inflammatory marker concentrations as well as further limit misclassification of SEP categories. Furthermore, education and occupation exposures were categorized as three-level variables, consequently socioeconomic misclassification is expected. In addition, for the measure of own occupation, the minimum age used for occupation assessment was 28 years in order to allow at least 10 years from the typical age of high school graduation to complete education, and obtain an occupation. Most participants were middle-aged or older when occupation was assessed. There was an approximately 40-year age range in participants at each measurement time, and approximately 20 years between Examinations 2 and 7 when occupation was assessed. Ideally all participants would be of similar ages when occupation was assessed, however these data were not available in this study. Given that occupation was categorized using only two or three broad categories, and that the mean age of assessment for the exam when participants were youngest (Examination 2) was 47.7 years, we felt this was a reasonable approximation of occupation. Misclassification of occupation is expected in this study, and other studies with occupation measured at narrower aged ranges will provide more precise information on the role of occupation in relation to inflammatory markers. Finally, the primary analyses represented potential risk for multiple statistical testing where there were 64 individual point estimates presented for age- and sex-adjusted regression analyses (Tables 2–4). Given an alpha of 0.05 used for the 95 percent confidence intervals, we would expect the reference categories' point estimates to be outside of comparison groups point estimates' 95 percent confidence intervals for 3 of the 64 tests simply due to chance. Given that the analyses demonstrated 35 of the 64 tests to have a reference category point estimates outside of the comparison groups point estimates' 95% confidence intervals, we feel that the overarching findings in this report are consistent with a relation between SEP and inflammation.

Conclusion

This study provides evidence that cumulative SEP across the life course is inversely associated with several inflammatory markers including CRP, IL-6, ICAM-1, TNFR2, LP-PLA_2 activity, MCP-1 and fibrinogen in age- and sex-adjusted analyses. Own education was associated with almost all studied inflammatory markers (CRP, IL-6, ICAM-1, TNFR2, P-selectin, MCP-1 and fibrinogen), while father's education, father's occupation and own occupation were associated with several but not all inflammatory markers in age- and sex-adjusted analyses (father's education: CRP, ICAM-1, TNFR2, Lp-PLA_2 activity; father's occupation: CRP, IL-6, ICAM-1, TNFR2 and MCP-1; offspring occupation: CRP, IL-6, ICAM-1 and TNFR2). There were few consistent trends between social mobility trajectories and most inflammatory markers, with general evidence that low SEP at any time in the life course conferred risk for elevated inflammatory markers. The strengths of association between the SEP measures and inflammatory markers were typically substantially accounted for by CHD risk markers, suggesting these may be important mechanisms that explain a reasonable amount of the association between SEP and the studied inflammatory markers. These data provide potential biological mechanistic evidence of inverse associations between life course SEP and CHD found in observational studies.

Appendix. Supplementary data

The supplementary data associated with this article can be found in the on-line version at doi:10.1016/j.socscimed.2010.03.012.

References

Almeida, O. P., Norman, P. E., Allcock, R., van Bockxmeer, F., Hankey, G. J., Jamrozik, K., et al. (2009). Polymorphisms of the CRP gene inhibit inflammatory response and increase susceptibility to depression: the Health in Men Study. *International Journal of Epidemiology, 38*(4), 1049–1059.

Armstrong, E. J., Morrow, D. A., & Sabatine, M. S. (2006). Inflammatory biomarkers in acute coronary syndromes: part IV: matrix metalloproteinases and biomarkers of platelet activation. *Circulation, 113*(9), e382–e385.

Borrell, L. N., & Crawford, N. D. (2008). Social disparities in periodontitis among United States adults 1999–2004. *Community Dentistry and Oral Epidemiology, 36*(5), 383–391.

Brunner, E., Davey Smith, G., Marmot, M., Canner, R., Beksinska, M., & O'Brien, J. (1996). Childhood social circumstances and psychosocial and behavioural factors as determinants of plasma fibrinogen. *Lancet, 347*(9007), 1008–1013.

Cesari, M., Penninx, B. W., Newman, A. B., Kritchevsky, S. B., Nicklas, B. J., Sutton-Tyrrell, K., et al. (2003). Inflammatory markers and onset of cardiovascular events: results from the Health ABC study. *Circulation, 108*(19), 2317–2322.

Colhoun, H. M., Hemingway, H., & Poulter, N. R. (1998). Socio-economic status and blood pressure: an overview analysis. *Journal of Human Hypertension, 12*(2), 91–110.

Danesh, J., Wheeler, J., Hirschfield, G., Eda, S., Eiriksdottir, G., Rumley, A., et al. (2004). C-reactive protein and other circulating markers of inflammation in the prediction of coronary heart disease. *The New England Journal of Medicine, 350*(14), 1387–1397.

Dantzer, R., O'Connor, J. C., Freund, G. G., Johnson, R. W., & Kelley, K. W. (2008). From inflammation to sickness and depression: when the immune system subjugates the brain. *Nature Reviews. Neuroscience, 9*(1), 46–56.

Elliott, P., Chambers, J. C., Zhang, W., Clarke, R., Hopewell, J. C., Peden, J. F., et al. (2009). Genetic loci associated with C-reactive protein levels and risk of coronary heart disease. *The Journal of the American Medical Association, 302*(1), 37–48.

Empana, J. P., Sykes, D. H., Luc, G., Juhan-Vague, I., Arveiler, D., Ferrieres, J., et al. (2005). Contributions of depressive mood and circulating inflammatory markers to coronary heart disease in healthy European men: the Prospective Epidemiological Study of Myocardial Infarction (PRIME). *Circulation, 111*(18), 2299–2305.

Galobardes, B., Shaw, M., Lawlor, D. A., Lynch, J. W., & Davey Smith, G. (2006). Indicators of socioeconomic position (part 1). *Journal of Epidemiology and Community Health, 60*(1), 7–12.

Galobardes, B., Smith, G. D., & Lynch, J. W. (2006). Systematic review of the influence of childhood socioeconomic circumstances on risk for cardiovascular disease in adulthood. *Annals of Epidemiology, 16*(2), 91–104.

Garza, C. A., Montori, V. M., McConnell, J. P., Somers, V. K., Kullo, I. J., & Lopez-Jimenez, F. (2007). Association between lipoprotein-associated phospholipase A2 and cardiovascular disease: a systematic review. *Mayo Clinic Proceedings, 82*(2), 159–165.

Gilman, S. E., Abrams, D. B., & Buka, S. L. (2003). Socioeconomic status over the life course and stages of cigarette use: initiation, regular use, and cessation. *Journal of Epidemiology and Community Health, 57*(10), 802–808.

Gilman, S. E., Martin, L. T., Abrams, D. B., Kawachi, I., Kubzansky, L., Loucks, E. B., et al. (2008). Educational attainment and cigarette smoking: a causal association? *International Journal of Epidemiology*.

Gimeno, D., Brunner, E. J., Lowe, G. D., Rumley, A., Marmot, M. G., & Ferrie, J. E. (2007). Adult socioeconomic position, C-reactive protein and interleukin-6 in the Whitehall II prospective study. *European Journal of Epidemiology, 22*(10), 675–683.

Gimeno, D., Ferrie, J. E., Elovainio, M., Pulkki-Raback, L., Keltikangas-Jarvinen, L., Eklund, C., et al. (2008). When do social inequalities in C-reactive protein start? A life course perspective from conception to adulthood in the Cardiovascular Risk in Young Finns Study. *International Journal of Epidemiology, 37*(2), 290–298.

Gonzalez, M. A., Rodriguez Artalejo, F., & Calero, J. R. (1998). Relationship between socioeconomic status and ischaemic heart disease in cohort and case-control studies: 1960–1993. *International Journal of Epidemiology, 27*(3), 350–358.

Hallqvist, J., Lynch, J., Bartley, M., Lang, T., & Blane, D. (2004). Can we disentangle life course processes of accumulation, critical period and social mobility? An analysis of disadvantaged socio-economic positions and myocardial infarction in the Stockholm Heart Epidemiology Program. *Social Science & Medicine, 58*(8), 1555–1562.

Kannel, W. B., Feinleib, M., McNamara, P. M., Garrison, R. J., & Castelli, W. P. (1979). An investigation of coronary heart disease in families. The Framingham Offspring study. *American Journal of Epidemiology, 110*(3), 281–290.

Keaney, J. F., Jr., Larson, M. G., Vasan, R. S., Wilson, P. W., Lipinska, I., Corey, D., et al. (2003). Obesity and systemic oxidative stress: clinical correlates of oxidative stress in the Framingham Study. *Arteriosclerosis, Thrombosis, and Vascular Biology, 23*(3), 434–439.

Keaney, J. F., Jr., Massaro, J. M., Larson, M. G., Vasan, R. S., Wilson, P. W., Lipinska, I., et al. (2004). Heritability and correlates of intercellular adhesion molecule-1 in the Framingham Offspring Study. *Journal of the American College of Cardiology, 44*(1), 168–173.

Keavney, B., Danesh, J., Parish, S., Palmer, A., Clark, S., Youngman, L., et al. (2006). Fibrinogen and coronary heart disease: test of causality by 'Mendelian randomization'. *International Journal of Epidemiology, 35*(4), 935–943.

Koster, A., Bosma, H., Penninx, B. W., Newman, A. B., Harris, T. B., van Eijk, J. T., et al. (2006). Association of inflammatory markers with socioeconomic status. *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences, 61*(3), 284–290.

Kuh, D., & Ben-Shlomo, Y. (2004). Introduction: a life course approach to the aetiology of adult chronic disease. In D. Kuh, & Y. Ben-Shlomo (Eds.), *A life course approach to chronic disease epidemiology* (2nd ed.). (pp. 3–14) New York: Oxford University Press.

Kuh, D., Ben-Shlomo, Y., Lynch, J., Hallqvist, J., & Power, C. (2003). Life course epidemiology. *Journal of Epidemiology and Community Health, 57*(10), 778–783.

Lawlor, D. A., Batty, G. D., Morton, S. M., Clark, H., Macintyre, S., & Leon, D. A. (2005). Childhood socioeconomic position, educational attainment, and adult cardiovascular risk factors: the Aberdeen children of the 1950s cohort study. *American Journal of Public Health, 95*(7), 1245–1251.

Lawlor, D. A., Smith, G. D., Rumley, A., Lowe, G. D., & Ebrahim, S. (2005). Associations of fibrinogen and C-reactive protein with prevalent and incident coronary heart disease are attenuated by adjustment for confounding factors. British Women's Heart and Health Study. *Thrombosis and Haemostasis, 93*(5), 955–963.

de Lemos, J. A., Morrow, D. A., Sabatine, M. S., Murphy, S. A., Gibson, C. M., Antman, E. M., et al. (2003). Association between plasma levels of monocyte chemoattractant protein-1 and long-term clinical outcomes in patients with acute coronary syndromes. *Circulation, 107*(5), 690–695.

Libby, P. (2006). Inflammation and cardiovascular disease mechanisms. *The American Journal of Clinical Nutrition, 83*(2), 456S–460S.

Linden, G. J., McClean, K., Young, I., Evans, A., & Kee, F. (2008). Persistently raised C-reactive protein levels are associated with advanced periodontal disease. *Journal of Clinical Periodontology, 35*(9), 741–747.

Loucks, E. B., Sullivan, L. M., Hayes, L. J., D'Agostino, R. B., Sr., Larson, M. G., Vasan, R. S., et al. (2006). Association of educational level with inflammatory markers in the Framingham Offspring Study. *American Journal of Epidemiology, 163*(7), 622–628.

Malik, I., Danesh, J., Whincup, P., Bhatia, V., Papacosta, O., Walker, M., et al. (2001). Soluble adhesion molecules and prediction of coronary heart disease: a prospective study and meta-analysis. *Lancet, 358*(9286), 971–976.

Maty, S. C., Everson-Rose, S. A., Haan, M. N., Raghunathan, T. E., & Kaplan, G. A. (2005). Education, income, occupation, and the 34-year incidence (1965–99) of Type 2 diabetes in the Alameda County Study. *International Journal of Epidemiology, 34*(6), 1274–1281.

McLaren, L. (2007). Socioeconomic status and obesity. *Epidemiologic Reviews, 29*, 29–48.

McNamara, J. R., & Schaefer, E. J. (1987). Automated enzymatic standardized lipid analyses for plasma and lipoprotein fractions. *Clinica Chimica Acta, 166*(1), 1–8.

Mendall, M. A., Patel, P., Asante, M., Ballam, L., Morris, J., Strachan, D. P., et al. (1997). Relation of serum cytokine concentrations to cardiovascular risk factors and coronary heart disease. *Heart, 78*(3), 273–277.

Muntaner, C., Eaton, W. W., Miech, R., & O'Campo, P. (2004). Socioeconomic position and major mental disorders. *Epidemiologic Reviews, 26*, 53–62.

Nazmi, A., & Victora, C. G. (2007). Socioeconomic and racial/ethnic differentials of C-reactive protein levels: a systematic review of population-based studies. *BMC Public Health, 7*, 212.

Pai, J. K., Pischon, T., Ma, J., Manson, J. E., Hankinson, S. E., Joshipura, K., et al. (2004). Inflammatory markers and the risk of coronary heart disease in men and women. *The New England Journal of Medicine, 351*(25), 2599–2610.

Panagiotakos, D. B., Pitsavos, C., Manios, Y., Polychronopoulos, E., Chrysohoou, C. A., & Stefanadis, C. (2005). Socio-economic status in relation to risk factors associated with cardiovascular disease, in healthy individuals from the ATTICA study. *European Journal of Cardiovascular Prevention & Rehabilitation, 12*(1), 68–74.

Pollitt, R. A., Kaufman, J. S., Rose, K. M., Diez-Roux, A. V., Zeng, D., & Heiss, G. (2007). Early-life and adult socioeconomic status and inflammatory risk markers in adulthood. *European Journal of Epidemiology, 22*(1), 55–66.

Pollitt, R. A., Kaufman, J. S., Rose, K. M., Diez-Roux, A. V., Zeng, D., & Heiss, G. (2008). Cumulative life course and adult socioeconomic status and markers of inflammation in adulthood. *Journal of Epidemiology and Community Health, 62*(6), 484–491.

Ramsay, S., Lowe, G. D., Whincup, P. H., Rumley, A., Morris, R. W., & Wannamethee, S. G. (2008). Relationships of inflammatory and haemostatic markers with social class: results from a population-based study of older men. *Atherosclerosis, 197*(2), 654–661.

Ridker, P. M., Cannon, C. P., Morrow, D., Rifai, N., Rose, L. M., McCabe, C. H., et al. (2005). C-reactive protein levels and outcomes after statin therapy. *The New England Journal of Medicine, 352*(1), 20–28.

Senese, L. C., Almeida, N. D., Kittler Fath, A., Smith, B. T., & Loucks, E. B. (2009). Associations between childhood socioeconomic position and adulthood obesity. *Epidemiologic Reviews*.

Tabassum, F., Kumari, M., Rumley, A., Lowe, G., Power, C., & Strachan, D. P. (2008). Effects of socioeconomic position on inflammatory and hemostatic markers: a life-course analysis in the 1958 British Birth Cohort. *American Journal of Epidemiology*.

General Internal Medicine
Boston University School of Medicine
2010 Publications – M-Z

1. Mahajan VS, **Pace CA**, Jarolim P. Interpretation of HIV serologic testing results. Clinical Chemistry. 2010;56(10):1523–27.

2. **Manze M**, **Rose AJ**, Orner MB, **Berlowitz DR**, **Kressin NR**. Understanding racial disparities in treatment intensification for hypertension management. J Gen Intern Med. 2010 Apr;25(8):819-25. doi: 10.1007/s11606-010-1342-9. PMID: 20386998.

3. Markuns JF, Fraser B, **Orlander JD**. The path to physician leadership in community health centers: Implications for training. Fam Med. 2010 Jun;42(6):403-7. PMID: 20526907.

4. McInnes DK, **Gifford AL**, Kazis LE, Wagner TH. Disparities in health-related internet use by US veterans: Results from a national survey. Inform Prim Care. 2010;18(1):59-68. PMID: 20429979.

5. McInnes DK, Hyun JK, Trafton JA, Asch SM, **Gifford AL**. Program characteristics associated with testing for HIV and hepatitis C in veterans substance use disorder clinics. Psychiatr Serv. 2010 Jan;61(1):90-4. PMID: 20044426.

6. **Medow MA**, Arkes HR, Shaffer VA. Are residents' decisions influenced more by a decision aid or a specialist's opinion? A randomized controlled trial. J Gen Intern Med. 2010 Apr;25(4):316-20. PMID: 20119873.

7. Middleton JC, Hahn RA, Kuzara JL, Elder R, Brewer R, Chattopadhyay S, Fielding J, **Naimi TS**, Toomey T, Lawrence B; 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 Dec;39(6):575-89. PMID: 21084079.

8. Mitchell SE, **Paasche-Orlow MK**, Forsythe SR, Chetty VK, O'Donnell JK, Greenwald JL, Culpepper L, Jack BW. Post-discharge hospital utilization among adult medical inpatients with depressive symptoms. J Hosp Med. 2010 Sep;5(7):378-84. PMID: 20577971.

9. Mostow C, **Crosson J**, **Gordon S**, **Chapman S**, Gonzalez P, Hardt E, **Delgado L**, James T, **David M**. Treating and precepting with RESPECT: A relational model addressing race, ethnicity and culture in medical training. J Gen Intern Med. 2010 May;25 Suppl 2:S146-54. doi:10.1007/s11606-010-1274-4. PMID: 20352510.

10. **Naimi T**. Commentary on McCaul et al. (2010): Observational studies about average alcohol consumption and health--closing time for a limited evidence base. Addiction. 2010 Aug;105(8):1401-2. PMID: 20653620.

11. **Naimi TS**, Nelson DE, Brewer RD. The intensity of binge alcohol consumption among U.S. adults. Am J Prev Med. 2010;38(2):201-7. PMID: 20626370.

12. Nazi KM, Hogan TP, Wagner TH, McInnes DK, Smith BM, Haggstrom D, Chumbler NR, **Gifford AL**, Charters KG, Saleem JJ, Weingardt KR, Fischetti LF, Weaver FM. Embracing a health services research perspective on personal health records: Lessons learned from the VA My HealtheVet system. J Gen Intern Med. 2010 Jan;25 Suppl 1:62-7. PMID: 20077154.

13. Nelson DE, **Naimi TS**, Brewer RD, Roeber J. US state alcohol sales compared to survey data, 1993-2006. Addiction. 2010 Sep;105(9):1589-96. PMID: 20626370.

14. Newman AB, Walter S, Lunetta KL, Garcia ME, Slagboom PE, Christensen K, Arnold AM, Aspelund T, Aulchenko YS, Benjamin EJ, Christiansen L, D'Agostino RB Sr, Fitzpatrick AL, Franceschini N, Glazer NL, Gudnason V, Hofman A, Kaplan R, Karasik D, Kelly-Hayes M, Kiel DP, Launer LJ, Marciante KD, Massaro JM, Miljkovic I, Nalls MA, Hernandez D, Psaty BM, Rivadeneira F, Rotter J, Seshadri S, Smith AV, Taylor KD, Tiemeier H, Uh HW, Uitterlinden AG, Vaupel JW, Walston J, Westendorp RGJ, Harris TB, Lumley T, van Duijn CM, **Murabito JM**. 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. 2010;1-10. doi:10.1093/gerona/glq028. PMID: 20304771. PMCID: PMC2854887.

15. **Paasche-Orlow MK**, **Gray W**, Cohen MJ. Crossword puzzle: A decoding challenge. J of Health Communication. 2010;15:224-25.

16. **Paasche-Orlow MK**, Wilson EAH, McCormack L. The evolving field of health literacy research. J of Health Communication. 2010;15:5-8. PMID: 20845188.

17. **Paasche-Orlow MK**, Wolf MS. Promoting health literacy research to reduce health disparities. J of Health Communication. 2010;15:34-41. PMID: 20845191.

18. Parker VA, Clark JA, Leyson J, Calhoun E, Carroll JK, **Freund KM**, **Battaglia TA**. Patient navigation: Development of a protocol for describing what navigators do. Health Research and Educational Trust. 2010 Apr;45(2):514-31. doi: 10.1111/j.1475-6773.2009.01079.x. PMID: 20132342.

19. Peköz EA, Shwartz M, Christiansen C, **Berlowitz D**. Approximate models for aggregate data when individual-level data sets are very large or unavailable. Statistics in Medicine. 2010;29:2180-93.

20. Perna FM, Craft L, **Freund KM**, Skrinar G, Stone M, Kachnic L, Youren C, **Battaglia TA**. The effect of a cognitive behavioral exercise intervention on clinical depression in a multi-ethnic sample of women with breast cancer. International Journal of Exercise and Sport Psychology. 2010;8(1):36-47.

21. Polat J, Feinberg E, **Crosby SS**. Ocular manifestations of torture: Solar retinopathy as a result of forced solar gazing. Br J Ophthalmol. 2010 Oct;94(10):1406-7. PMID: 20576780.

22. Pololi L, Cooper LA, **Carr P**. Race, disadvantage and faculty experiences in academic medicine. J Gen Intern Med. 2010;25(12):1363-69. PMID: 20697960.

23. Pugh MJV, VanCott AC, Steinman MA, Mortensen EM, Amuam ME, Wang C-P, Knoefel JE, **Berlowitz DR**. Choice of initial antiepileptic drug for older veterans: Possible pharmacokinetic drug interactions with existing medications. J Am Geriatr Soc. 2010;58(3):465-71. PMID: 20398114.

24. **Ramani S**, Ring BN, Lowe R, Hunter D. A Pilot study assessing knowledge of clinical signs and physical examination skills in incoming medicine residents. Journal of Graduate Medical Education. 2010 Jun;2(2):232-35.

25. **Rose AJ**, **Hylek EM**. High blood pressure while taking antithrombotic medication is associated with an increased risk of developing intracranial haemorrhage. Evid Based Med. 2010 Dec;15(6):189-90. PMID: 20797990.

26. **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 Oct;8(10):2182-91. doi: 10.1111/j.1538-7836.2010.03996.x. PMID: 20653840.

27. **Rose AJ**, Ozonoff A, **Henault LE**, **Hylek EM**. Anticoagulation for valvular heart disease in community-based practice. Thromb Haemost. 2010 Feb;103(2):329-37. PMID: 20024499.

28. Rosenquist JN, **Murabito J**, Fowler JH, Christakis NA. The spread of alcohol consumption behavior in a large social network. Ann Intern Med. 2010;152(7):426-33.

29. **Saitz R**. Candidate performance measures for screening for, assessing, and treating unhealthy substance use in hospitals: Advocacy or evidence-based practice? Ann Intern Med. 2010 Jul;153(1):40-3. PMID: 20621901.

30. **Saitz R**. Alcohol screening and brief intervention in primary care: Absence of evidence for efficacy in people with dependence or very heavy drinking. Drug Alcohol Rev. 2010;29(6):631-40. PMID: 20973848. PMCID: PMC2966031.

31. **Saitz R**. Evidence-based medicine: Time for transition and translation (to practice). Evid Based Med. 2010 Aug;15(4):103-4. PMID: 20570949.

32. **Saitz R**. Most inpatients with unhealthy alcohol use have an alcohol use disorder. Int J Public Health. 2010 Jun;55(6):527-28. PMID: 20535524.

33. **Saitz R**, **Alford DP**, Bernstein J, **Cheng DM**, **Samet J**, Palfai T. Screening and brief intervention for unhealthy drug use in primary care settings: Randomized clinical trials are needed. J Addict Med. 2010;4(3):123-30. PMID: 20936079. PMCID: PMC2950314.

34. **Saitz R**, **Naimi TS**. Adolescent alcohol use and violence: Are brief interventions the answer? JAMA. 2010 Aug;304(5):575-7. PMID: 20682942.

35. **Samet JH, Walley AY.** Interventions targeting HIV infected risky drinkers: Drops in a bottle. Alcohol Research & Health. 2010;33(3):267-79.

36. **Samet JH**, **Pace CA**, **Cheng DM**, Coleman S, **Bridden C**, Pardesi M, Saggurti N, **Raj A**. Alcohol use and sex risk behaviors among HIV-infected female sex workers (FSWs) and HIV-infected male clients of FSWs in India. AIDS Behav. 2010;14:S74-S83. PMID: 20544381.

37. Schmitt B, Steendijk P, Ovroutski S, **Lunze K**, Rahmanzadeh P, Maarouf N, Ewert P, Berger F, Kuehne T. Pulmonary vascular resistance, collateral flow and ventricular function in patients with a Fontan Circulation at rest and during dobutamine stress. Circ Cardiovasc Imaging. 2010 Sep;3(5):623-31. PMID: 20631032.

38. Schwartz S, **Hoyte J**, James T, Conoscenti L, Johnson R, **Liebschutz J**. Challenges to engaging black male victims of community violence in healthcare research: Lessons learned from two studies. Psychological Trauma: Theory, Research, Practice and Policy. 2010;2(1):54-62. NIHMS197969.

39. Selim AJ, **Berlowitz D**, Kazis LE, Rogers W, Wright SM, Qian SX, Rothendler JA, Spiro A 3rd, Miller D, **Selim BJ**, Fincke BG. Comparison of health outcomes for male seniors in the Veterans Health Administration and Medicare Advantage plans. Health Serv Res. 2010 Apr;45(2):376-96. PMID: 20050934.

40. Sengupta S, Lo B, Strauss RP, Eron J, **Gifford AL**. How researchers define vulnerable populations in HIV/AIDS clinical trials. AIDS and Behavior. 2010;14(6):1313-19. PMID: 20721614. PMCID: PMC2975789.

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 Mar;25(8):803-8. doi: 10.1007/s11606-010-1311-3. PMID: 20237960. PMCID: PMC2896583.

42. Shwartz M, Burgess JF, **Berlowitz D**. Benefit-of-the-doubt approaches for calculating a composite measure of quality. Health Services and Outcomes Research Methodology. 2010;9(4):234-51.

43. **Smith PC**, Schmidt SM, Allensworth-Davies D, **Saitz R**. A single-question screening test for drug use in primary care. Arch Intern Med. 2010 Jul;170(13):1155-60.PMID: 20625025.

44. Squires LE, **Alford DP**, Bernstein J, Palfai T, **Saitz R**. Screening and brief intervention for drug use in primary care. J Addict Med. 2010;4(3):131-36.

45. 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 Jun;152(11):712-20. PMID: 20513829.

46. **Stillman M**. Concierge medicine: A "regular" physician's perspective. Ann Intern Med. 2010 Mar;152(6):391-92.

47. **Stillman M**. For those on whom we rely. JAMA. 2010;303(15):1458-59.

48. Taft C, Schwartz S, **Liebschutz JM**. Intimate partner aggression perpetration in primary care chronic pain patients. Victims and Violence. 2010;25(5):649-61. PMID: 21061870.

49. Tokuda Y, Okubo T, Yanai H, Doba N, **Paasche-Orlow MK**. Development and validation of a 15-item Japanese health knowledge test. J Epidemiol 2010;20(4):319-328. PMID: 20551582.

50. **Tsui JI**, Herman DS, Kettavong M, **Alford D**, Anderson BJ, Stein MD. Physician introduction to opioids for pain among patients with opioid dependence and depressive symptoms. J Subst Abuse Treat. 2010 Dec;39(4):378-83. PMID: 20727704.

51. Volandes AE, Barry MJ, Chang Y, **Paasche-Orlow MK**. Improving decision making at the end of life with video images. Med Decis Making. Jan-Feb 2010;30(1):29-34. PMID: 19675323.

52. Williams EC, Palfai T, **Cheng DM**, **Samet JH**, Bradley KA, Koepsell TD, Wickizer TM, Heagerty PJ, **Saitz R**. Physical health and drinking among medical inpatients with unhealthy alcohol use: A prospective study. Alc Clin Exp Res. 2010;34(7):1257-65. PMID: 20477765.
53. 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 Apr;8(4):744-9. PMID: 20398186.

54. Zallman L, Ma J, Xiao L, **Lasser KE**. Quality of US primary care delivered by resident and staff physicians. J Gen Intern Med. 2010 Nov;25(11):1193-7. doi:10.1007/s11606-010-1456-0. PMID: 20645018. PMCID: PMC2947643.

55. Zúñiga ML, Blanco E, Palinkas L, Strathdee SA, **Gifford AL**. Cross-cultural considerations in the recruitment of Latinos of Mexican-origin into HIV/AIDS clinical trials in the U.S.-Mexico border region: Clinician and patient perspectives. Journal of Immigrant and Refugee Studies. 2010;8(3):241-60.

Clinical Chemistry 56:10
1523–1527 (2010)

Clinical Case Study

Interpretation of HIV Serologic Testing Results

Vinay S. Mahajan,[1] Christine A. Pace,[2] and Petr Jarolim[1*]

CASE

A 33-year-old male patient visited the outpatient clinic at Brigham and Women's Hospital for a routine follow-up for obesity, obstructive sleep apnea, allergic rhinitis, and depression. He was maintained on a nocturnal continuous positive airway pressure device, loratadine, duloxetine, and fluticasone nasal spray. He was a resident of Boston and had not traveled outside the country. He denied intravenous drug use or high-risk sexual behavior, and he had not received any blood products. He had received his most recent influenza vaccine about 6 months earlier. He was screened for type 2 diabetes and hyperlipidemia. As a part of routine clinical care, he was also offered HIV screening in accordance with the current CDC recommendations *(1)*. The HIV assay [HIV 1/O/2 Enhanced (EHIV)], which was performed on the ADVIA Centaur analyzer (Siemens Healthcare Diagnostics), yielded a reactive result. As per the assay protocol developed by the manufacturer, the initially reactive sample was retested in duplicate after centrifugation; both results were reactive. The positive screen was followed up with a confirmatory western blot (WB)[3] analysis, which yielded an indeterminate result. The presence of an isolated p24 band in the WB (GS Western HIV-1; Bio-Rad Laboratories) was of concern regarding possible early HIV seroconversion.

QUESTIONS TO CONSIDER

1. What factors are known to cause false-positive HIV serologic test results?
2. What factors are known to cause an indeterminate WB result?
3. What further testing or clinical history would be of help in evaluating a patient with an indeterminate WB result?

[1] Departments of Pathology and [2] Internal Medicine, Brigham and Women's Hospital, Boston, MA.
* Address correspondence to this author at: Department of Pathology, Brigham and Women's Hospital, 75 Francis St., Boston, MA 02115. Fax 617-731-4872; e-mail pjarolim@partners.org.
Received October 26, 2009; accepted March 22, 2010.
DOI: 10.1373/clinchem.2009.139535
[3] Nonstandard abbreviations: WB, western blot; NAT, nucleic acid testing.

DISCUSSION

The Siemens EHIV screen performed in this case is a double antigen-bridging microparticle chemiluminometric immunoassay that detects antibodies against p24, gp41, gp120 (from HIV-1), gp36 (from HIV-2), and a synthetic peptide from group O HIV-1 (Fig. 1). A positive result indicates the presence of antibodies that recognize any of these antigens, regardless of their isotype or subclass. Although such third-generation HIV immunoassays have greatly improved analytical sensitivity and specificity, false-positive results have not been eliminated completely. A common cause of false-positive serologic screens for HIV is recent influenza vaccination or an incidental viral infection *(2, 3)*. In addition to flu vaccination and viral infections, false-positive HIV-1 immunoassay results have been reported in a variety of other conditions, such as autoimmune disease, renal failure, cystic fibrosis, multiple pregnancies, blood transfusions, liver diseases, parenteral substance abuse, hemodialysis, and vaccinations against hepatitis B and rabies *(4)*. Thus, a positive result in an HIV screening test must be followed up with a more specific confirmatory test.

WB is routinely used to confirm a reactive HIV serologic screening result. These assays, which contain either viral lysate or recombinant HIV proteins, allow the determination of the antigenic specificity of the antibodies in the patient's serum. The predominant type of HIV in the US is HIV-1. A confirmatory test for HIV-1 infection was recommended because this patient had not traveled to any part of the world with a high prevalence of HIV-2, such as West Africa. The major antibody specificities detected in HIV-1 WB analysis include gp160, gp120, p65, p55, gp41, p40, p31, and p24. To be reported as positive, the WB assay requires reactivity against the gp41 and gp120/160 bands encoded by the *env*[4] (gp160, envelope glycoprotein) gene or against either one of these *env* bands plus the p24 band encoded by *gag* [Pr55(Gag)]. Such a result is highly specific for the presence of HIV infection *(5)*. A negative result implies the absence of any of the above bands. The result is called indeterminate when the band profile does not meet the criteria for a positive

[4] Genes: *env*, gp160, envelope glycoprotein [HIV-1 gene]; *gag*, Pr55(Gag) [HIV-1 gene].

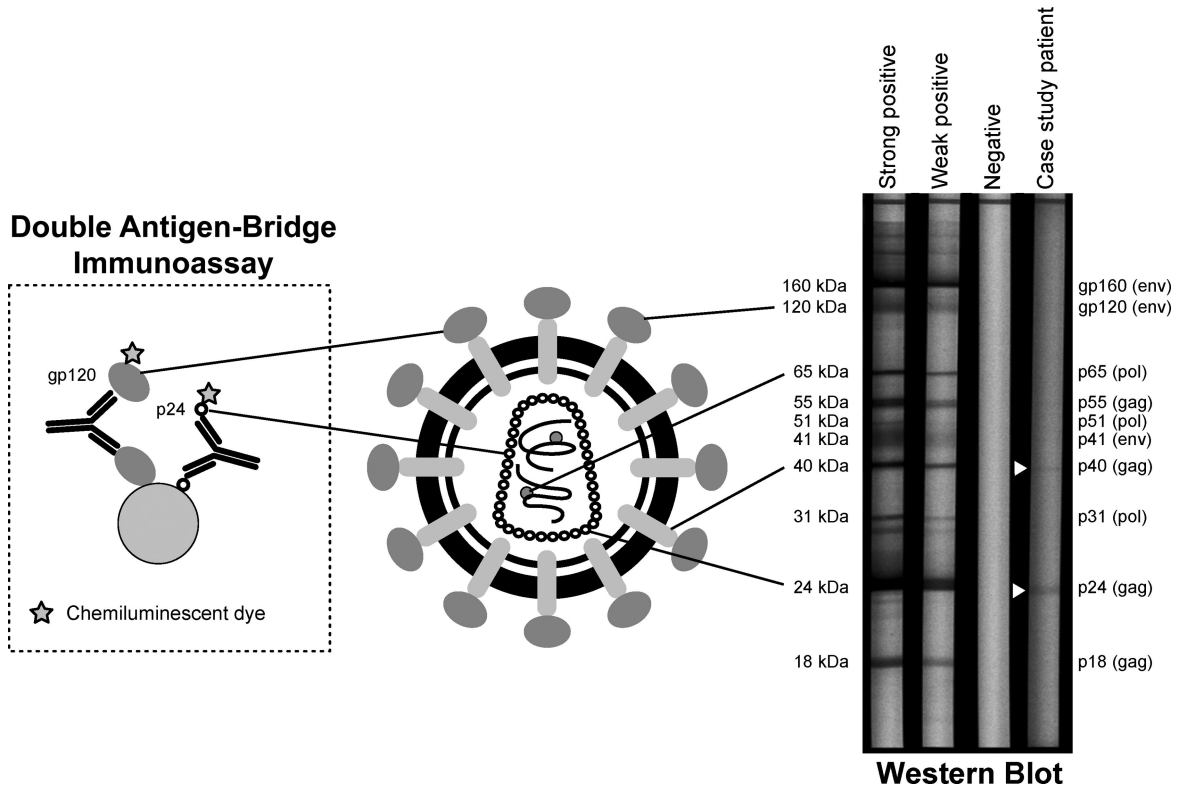


Fig. 1. Serologic tests for HIV.

The diagnosis of HIV infection is based on the detection of HIV antibodies against multiple HIV antigens. Antibodies against HIV are screened with an assay such as a double antigen-bridge microparticle immunoassay (left) and individually confirmed with a WB (right). The HIV WB from the patient in this case study is shown in the far right lane.

result. The patient's WB yielded an indeterminate result. In this case, the result was reported as indeterminate because a sharp p24 band and a weak p40 band were observed (Fig. 1).

After exposure to HIV, it usually takes an individual at least 3 weeks to build up an antibody titer sufficient to be detectable by a third-generation HIV immunoassay. This period is called the "seroconversion window." Because antibodies to p24 develop early during the course of infection, an indeterminate WB pattern seen during this window is often associated with an isolated p24 band *(6)*. Qualitative reverse-transcription PCR analysis for HIV is used to screen for or confirm the presence of HIV infection during the seroconversion window, and the screen can become positive as early as 10 days after exposure *(7)*. Nucleic acid testing (NAT) for HIV is also used when a rare HIV genotype is suspected; such testing plays a critical role in neonatal HIV screening, owing to interference from maternal antibodies. Besides early seroconversion, other causes of indeterminate HIV-1 WB results in the setting of HIV infection include infection with HIV-2 and advanced AIDS *(6)*. An indeterminate WB result can also arise from antibodies that are cross-reactive to HIV antigens, such as those associated with HTLV infection; with vaccination against influenza, hepatitis, or rabies; or with animal handlers exposed to unusual viruses. Nonspecific antibody binding to nonviral cellular components in the WB HIV antigen preparation can also produce an indeterminate WB result. Such results may be associated with frequent transfusions, injection drug use, liver disease, multiple pregnancies, rheumatoid factor, lymphoma, multiple sclerosis, various autoimmune disorders, a positive result in the rapid plasma reagin test, and chronic hemodialysis *(6)*.

The patient was contacted for follow-up of his HIV test results and possible NAT. Upon further questioning, however, he recalled that he had received an experimental HIV vaccine >5 years earlier. HIV vaccines may include either *gag*- or *env*-encoded proteins or both. Vaccines designed to induce cell-mediated immunity can also elicit a humoral response and produce vaccine-induced seropositivity. A majority of *gag* vac-

cine recipients have p24, p40, and/or p55 bands in their WB *(8)*. *env* vaccine recipients can have gp41, gp120, and gp160 bands. Such WB results are often reported as indeterminate, but some HIV vaccine recipients can meet the criteria for a positive HIV WB result. These patients can pose a true diagnostic challenge.

The results of HIV testing of vaccine recipients can be easily misinterpreted and can have a negative social impact *(9)*. Because of the blinding procedures of many vaccine trial designs, neither the patients nor the researchers may know whether a placebo or an experimental vaccine was administered. Vaccine-induced seropositivity can potentially lead to unblinding of the study participants as well as researchers, with a risk of compromising the study data. Therefore, HIV testing of vaccine trial participants is usually performed in designated laboratories with appropriate anonymization protocols that can provide interpretation of results without the risk of unblinding. The results of HIV testing in vaccine recipients need to be confirmed with NAT.

Vaccine trial participants are counseled to undergo HIV testing exclusively with the vaccine research group. Follow-up periods in such trials extend for decades, however, and patients may not recall all of the details. Therefore, the National Institute of Allergy and Infectious Diseases has provided participants in NIH-supported HIV vaccine trials with both a toll-free number for assistance and identification cards that document study participation *(9)*. A large number of experimental HIV vaccine trials have been undertaken over the last 2 decades, and there is a steadily increasing population of recipients of experimental HIV vaccines who present for HIV screening. This trend is likely to continue, especially considering the encouraging results of the recent HIV vaccine trial in Thailand *(10)*. Misinterpretation of the results of off-site HIV tests in vaccine trial volunteers may best be avoided through better communication between HIV vaccine researchers and local providers of diagnostic tests.

The decision for confirmatory NAT was deferred, and the vaccine research group was notified for appropriate interpretation, follow-up, and counseling regarding the patient's HIV screening result, in accordance with the study protocol. This procedure ensured that both the patient and the researchers remained blinded to whether the patient received a placebo or a test dose of the experimental vaccine.

POINTS TO REMEMBER

- False-positive HIV serologic screens can be caused by recent influenza vaccination, incidental viral infections, autoimmune disease, renal failure, cystic fibrosis, multiple pregnancies, blood transfusions, liver diseases, parenteral substance abuse, hemodialysis, or vaccinations against hepatitis B and rabies.
- An indeterminate WB result can be caused by a weak titer of anti–HIV-1 antibodies (as seen in early seroconversion), advanced AIDS, infection with an unusual HIV type, or recipients of experimental HIV vaccines. It can also be caused by the presence of antibodies cross-reactive against HIV antigens (incidental viral infection; vaccination against influenza, hepatitis, or rabies; or HTLV infection) or reactivity to the nonviral components of the WB (various autoimmune disorders, multiple pregnancies, and recipients of multiple blood transfusions).
- An indeterminate WB result should be followed up with qualitative NAT if early seroconversion is suspected, with a repeat immunoassay and WB analysis performed in 2–4 weeks. Although the US Food and Drug Administration has not cleared the use of quantitative viral load for HIV diagnosis, the viral load is unlikely to be $<$5000 copies/mL during acute HIV infection. Persistent reactivity of the antibody screening assay with a simultaneous lack of any change in the WB pattern suggests the absence of HIV infection.
- An increasing number of recipients of experimental HIV vaccines, which can cause false-positive results in HIV serologic tests, are being offered HIV screening. Whenever possible, testing for HIV in such patients is best performed in consultation with the vaccine research group responsible for the trial. This procedure will ensure proper interpretation of test results without compromising the study data.

Author Contributions: *All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.*

Authors' Disclosures of Potential Conflicts of Interest: *Upon manuscript submission, all authors completed the Disclosures of Potential Conflict of Interest form. Potential conflicts of interest:*

Employment or Leadership: None declared.
Consultant or Advisory Role: P. Jarolim, Gerson Lehrman Group and T2 Biosystems.
Stock Ownership: None declared.
Honoraria: P. Jarolim, Siemens Healthcare Diagnostics and Ortho Clinical Diagnostics.
Research Funding: P. Jarolim, Siemens Healthcare Diagnostics and Roche Diagnostics.
Expert Testimony: None declared.

Role of Sponsor: The funding organizations played no role in the design of study, choice of enrolled patients, review and interpretation of data, or preparation or approval of manuscript.

References

1. Branson BM, Handsfield HH, Lampe MA, Janssen RS, Taylor AW, Lyss SB, Clark JE. Revised recommendations for HIV testing of adults, adolescents, and pregnant women in health-care settings. MMWR Recomm Rep 2006; 55(RR-14):1–17.
2. Erickson CP, McNiff T, Klausner JD. Influenza vaccination and false positive HIV results. N Engl J Med 2006;354:1422–3.
3. Simonsen L, Buffington J, Shapiro CN, Holman RC, Strine TW, Grossman BJ, et al. Multiple false reactions in viral antibody screening assays after influenza vaccination. Am J Epidemiol 1995;141:1089–96.
4. Celum CL, Coombs RW, Jones M, Murphy V, Fisher L, Grant C, et al. Risk factors for repeatedly reactive HIV-1 EIA and indeterminate western blots. A population-based case-control study. Arch Intern Med 1994;154:1129–37.
5. Interpretive criteria used to report western blot results for HIV-1-antibody testing—United States. MMWR Morb Mortal Wkly Rep 1991;40:692–5.
6. Guan M. Frequency, causes, and new challenges of indeterminate results in Western blot confirmatory testing for antibodies to human immunodeficiency virus. Clin Vaccine Immunol 2007;14:649–59.
7. Feinberg MB. Changing the natural history of HIV disease. Lancet 1996;348: 239–46.
8. Quirk EK, Mogg R, Brown DD, Lally MA, Mehrotra DV, DiNubile MJ, Robertson MN. HIV seroconversion without infection after receipt of adenovirus-vectored HIV type 1 vaccine. Clin Infect Dis 2008;47:1593–9.
9. Allen M, Lau CY. Social impact of preventive HIV vaccine clinical trial participation: a model of prevention, assessment and intervention. Soc Sci Med 2008;66:945–51.
10. Rerks-Ngarm S, Pitisuttithum P, Nitayaphan S, Kaewkungwal J, Chiu J, Paris R, et al. Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand. N Engl J Med 2009;361:2209–20.

Commentary

Frederick S. Nolte*

This case report adds HIV vaccination to the list of well-known causes of false-positive results in HIV antibody–screening tests and illustrates the problems often associated with the interpretation of WBs. Timely and effective means of confirming HIV screening tests have become increasingly important as more centers integrate HIV screening into routine clinical care as recommended by the CDC.

In many laboratory settings, NATs for HIV-1 RNA are more widely available and less costly than WB and are not subject to indeterminate results. Although quantitative HIV-1 NATs have not been cleared by the US Food and Drug Administration (FDA) for diagnosis, they have been used for years for evaluating patients thought to be acutely infected. These patients typically have high viral loads ranging from 10^5 to 10^6 copies/mL, and the results present no problems with interpretation. The reports of false-positive results in viral load tests all occurred with a single method (Versant bDNA test; Siemens Healthcare Diagnostics); results were $<10^4$ copies/mL.

The APTIMA HIV-1 RNA Qualitative Assay (Gen-Probe), currently the only NAT that has been FDA-cleared for diagnosis of infection, can be used to diagnose neonatal and acute infections, confirm positive results in antibody-screening tests, and resolve indeterminate WB results. As the authors point out, a NAT was the next step in the diagnostic work-up, but it was deferred when the vaccination history was obtained.

The rare individuals who are infected with HIV-1 but who progress to AIDS either very slowly or not at all pose another diagnostic dilemma. In these long-term nonprogressors, HIV-1 antibody is easily demonstrated, but these individuals show low or undetectable HIV-1 RNA loads in the assays available to clinical laboratories. Viremic controllers have low but readily measurable virus loads. Elite controllers suppress HIV to extremely low concentrations, which are measurable only with the most analytically sensitive laboratory techniques.

Department of Pathology and Laboratory Medicine, Medical University of South Carolina, Charleston, SC.
* Address correspondence to the author at: Medical University of South Carolina, Pathology and Laboratory Medicine, 165 Ashley Ave., Suite 309, MSC 908, Charleston, SC 29425. Fax 843-792-7060; e-mail nolte@musc.edu.
Received July 23, 2010; accepted August 2, 2010.
DOI: 10.1373/clinchem.2010.153130

Author Contributions: *All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.*

Authors' Disclosures of Potential Conflicts of Interest: *Upon manuscript submission, all authors completed the Disclosures of Potential Conflict of Interest form. Potential conflicts of interest:*

Employment or Leadership: None declared.
Consultant or Advisory Role: F.S. Nolte, Gen-Probe and Abbott Molecular.
Stock Ownership: None declared.
Honoraria: None declared.
Research Funding: None declared.
Expert Testimony: None declared.

Role of Sponsor: The funding organizations played no role in the design of study, choice of enrolled patients, review and interpretation of data, or preparation or approval of manuscript.

Understanding Racial Disparities in Treatment Intensification for Hypertension Management

Meredith Manze, MPH[1,2], Adam J. Rose, MD, MSc[3,1], Michelle B. Orner, MPH[3], Dan R. Berlowitz, MD, MPH[3,2], and Nancy R. Kressin, PhD[4,3,1]

[1]Section of General Internal Medicine, Boston University School of Medicine, Boston, MA, USA; [2]Health Policy & Management Department, Boston University School of Public Health, Boston, MA, USA; [3]Center for Health Quality, Outcomes, and Economic Research, Bedford VA Medical Center, Bedford, MA, USA; [4]VA Boston Healthcare System, Boston, MA, USA.

BACKGROUND: Disparities in blood pressure (BP) control may be a function of disparities in treatment intensification (TI).

OBJECTIVE: To examine racial differences in TI, understand modifiable factors that may mediate this relationship, and explore the relative effects of TI and race on blood pressure.

DESIGN: Prospective cohort study.

PARTICIPANTS: Participants were 819 black and white patients with hypertension from an urban, safety-net hospital

MAIN MEASURES: We sequentially explored the effects of patient race, sociodemographic and clinical characteristics, beliefs about BP/medications, perceptions of provider/discrimination, sodium intake, medication adherence, and provider counseling on TI, performing a series of random effects analyses. To assess the effects of race and TI on BP, we performed linear regressions, using systolic BP (SBP) as the outcome.

KEY RESULTS: Unadjusted analyses and those including sociodemographic and clinical characteristics revealed that black patients had less TI than whites (−0.31 vs.−0.24, p<0.001), but adjustment for patient beliefs and experiences eliminated the effects of race (β= −0.02, p=0.5). Increased patient concerns about BP medications were related to lower TI, as was more provider counseling (β=−0.06, p=0.02 and β= −0.01, p=0.001, respectively). In the unadjusted analysis, black race was a significant predictor of SBP (134 mm/Hg for blacks vs. 131 mm/Hg for whites, p= 0.009), but when both race and TI were included in the model, TI was a significant predictor of SBP (final SBP 2.0 mm/Hg lower for each additional therapy increase per 10 visits, p<0.001), while race was not (Blacks 1.6 mm/Hg higher than whites, p=0.17).

CONCLUSIONS: Improved patient–provider communication targeted towards addressing patient concerns about medications may have the potential to reduce racial disparities in TI and ultimately, BP control.

Poster Presentation: AcademyHealth Annual Research Meeting, Chicago, IL, June 28–30th, 2009

Received September 23, 2009
Revised January 29, 2010
Accepted March 18, 2010
Published online April 13, 2010

KEY WORDS: disparities; treatment intensification; hypertension.
J Gen Intern Med 25(8):819-25
DOI: 10.1007/s11606-010-1342-9

BACKGROUND

Racial disparities in hypertension control and hypertension-related outcomes persist despite efforts to improve control and reduce disparities.[1–5] Approximately 34% of non-Hispanic blacks have hypertension, the highest prevalence of any racial/ethnic group.[5] Black individuals also have more severe hypertension with an earlier age of onset, compared to non-Hispanic whites.[6] For these reasons, blacks have an increased rate of adverse hypertension-related health outcomes and mortality.[6,7] Identifying contributing factors responsible for this disparity is crucial in improving health outcomes.

We have only a limited understanding of the reasons why blood pressure (BP) control is worse in blacks. Increasingly, clinical inertia, the phenomenon by which providers do not always initiate or intensify therapy in response to uncontrolled BP, is recognized as a major barrier to BP control.[8–10] Treatment intensification (TI) for hypertension occurs when providers initiate and intensify therapy for patients with elevated BP. TI has been linked to improved BP control.[9,11,12] TI is complex and often involves discussions between patients and providers. Patient beliefs about hypertension and their medications, as well as their adherence to antihypertensive medications and their experiences with care, are likely to be critical factors in these decisions to intensify therapy. TI is best viewed as something that patients and providers accomplish in collaboration. Therefore, addressing insufficient TI is likely to involve factors related to patient–provider communication, and patients' experience of the process of care, including perceptions and experiences of discrimination. A more complete understanding of relational determinants of TI may help in designing interventions to increase TI and thus improve BP control.

Few studies have examined the issue of racial disparities in TI, or whether this disparity may also explain racial disparities in BP control.[13,14] Therefore, our objectives are to 1) explore the extent of racial disparities in TI in hypertension care, 2)

elucidate the contributions of patient characteristics, beliefs and behaviors, and patient–provider interactions to racial disparities in TI, and 3) examine the relative effects of race and TI on BP control.

METHODS

Study sample

We identified all white and black patients ages 21 and older with primary care clinic visits at an urban safety-net hospital, diagnosis of hypertension, and prescribed at least one antihypertensive medication. (The term "black" includes patients of black race born in Africa, Caribbean or U.S.A.) We enrolled 869 patients, as part of a larger study, and collected baseline data on BP control and patient beliefs and attitudes about and experiences with BP care and medications. We then implemented an intervention where providers were randomized to receive an educational workshop aimed at improving communication about hypertension care. Patients were approached for participation during their regular clinic visit. After consenting, recruited participants completed a survey, and clinical information was abstracted from the medical record.[15] Of the patients enrolled in the parent study, 50 were excluded from the present analysis because they had ≤ 2 BP values, too few to characterize TI. Therefore, 819 patients with hypertension constituted our study population. This study was approved by our Institutional Review board.

Measures

Independent Variables.

Sociodemographic and Clinical Characteristics. Patient sociodemographic characteristics including race, education and income were obtained through self-report. Patients' clinical data was extracted from the electronic medical record (EMR), including age, gender, height, weight, number of BP medications and diagnosis of hypertension. The EMR was also used to obtain diagnoses of comorbid conditions which pertain to hypertension management, including benign prostatic hypertrophy, cerebrovascular disease, chronic kidney disease, congestive heart failure, coronary artery disease, diabetes mellitus, hyperlipidemia, nicotine dependence, obesity, and peripheral vascular disease.[6] A patient was considered obese if s/he had either a diagnosis of obesity in the EMR or a body mass index of at least 30.

Health Beliefs and Illness Perceptions. We examined patient beliefs and perceptions about high BP and antihypertensive medications, using the Beliefs about Medicines Questionnaire (BMQ-specific). The BMQ-specific includes ten items to evaluate patients' *concerns* about potential adverse effects from their BP medications and eight items to measure patients' beliefs regarding the *necessity* of their medications (five point scale ranging from Strongly Agree to Strongly Disagree).[16,17] Scores were summed within each of these scales to create an overall scale score ('necessity' scale alpha=0.81 and 'concerns' scale alpha=0.80). Each score was divided by the number of items to obtain a mean summary score, where a higher number indicated either greater concerns about or beliefs in the necessity of medications. Scale scores were created only if 75% of the items were answered.

To assess the degree of seriousness with which patients perceived hypertension and its sequellae, we utilized four additional items (Table 2; ranging from "extremely serious" to "not at all serious"). We used five separate dichotomous items to assess patients' beliefs about BP medications. These items were first created and utilized in our prior work with a similar patient population.[18]

We included ten items from the "cause" scale of the Illness Perception Questionnaire. These items were analyzed separately to examine patients' subjective beliefs about the etiology of their high BP (five point scale ranging from Strongly Agree to Strongly Disagree).[19]

Perceptions of Provider and Experiences of Discrimination. To assess patients' perceptions of their providers, we used three items from the Commonwealth Fund 2001 Health Care Quality Survey [20]. We created an additional question about the patients' perception of their providers' understanding of their cultural background and how it affects their health. Each item was scored individually.

To measure perceived discrimination in health care, we included the seven item measure developed by Bird and Bogart,[21] creating a dichotomous variable for anyone who answered 'yes' to any question compared to patients who responded 'no' to all questions.

Sodium Intake. Because dietary sodium is an important contributor to BP, we assessed patients' sodium intake using the three-item subscale within the previously validated Hill-Bone Compliance to High Blood Pressure Therapy Scale, with responses ranging on a four point Likert scale from "None of the time" to "All of the time",[22] summing the items to create one sodium intake score.

Medication Adherence. We assessed medication adherence because better adherence to antihypertensive medications is associated with improved BP control. Patients used an electronic recording device (MEMS cap), that recorded each time a patient opened the medication bottle to take his/her BP medication, for approximately 90 days following their enrollment into the study. We calculated the proportion of days in this period on which the patient took at least the number of doses prescribed. We used this data to categorize patients as having "poor adherence" (defined as less than 50% adherent), "fair adherence" (50–80% adherent), "excellent adherence" (greater than 80% adherent), or having missing MEMS cap data.

Provider Counseling. We assessed the content of the patient–provider discussion regarding hypertension care and management, following earlier work from Kressin and Pbert.[18,23] We included a series of 12 dichotomous items that assessed whether or not the provider asked or advised patients about various issues related to antihypertensive medication adherence.

We summed the items from this measure to create a summary scale score, where higher scores indicate more discussion of hypertension related issues. Scale scores were created only if 75% of the items were answered (alpha=0.86).

Dependent Variable.

Treatment Intensification. TI was our main dependent variable. One author (AJR) performed a manual chart review to measure TI scores. A subset of patients, representing 5% of all clinic visits, were randomly selected for blind reabstraction by another author (DRB). Agreement between the reviewers was good (k=0.93 (95% CI, 0.87 to 0.98)).[15] We used the following formula to measure TI: (visits with medication changes–visits with elevated BP) / number of clinic visits.[12,15] A TI score of zero signifies that treatment was intensified once for each visit with elevated BP. A score greater than zero signifies that treatment was intensified at more visits than there were visits with elevated BP, while a score less than zero signifies that there were more visits with elevated BP than episodes of TI. A unit of 0.1 on this scale indicates one more or less TI than expected per 10 visits. This definition of TI, known as the Standard-Based method, is the preferred measurement of TI in hypertension care.[15]

The expected number of medication increases was the number of occasions on which the recorded BP was elevated, defined as 140/90 mm/Hg or higher. BP values were taken from the medical record at the clinic visit. In prior work using this data set, we found that using the threshold of >= 130/80 for uncontrolled BP, for patients with diabetes or chronic kidney disease, yielded similar results as using the same threshold (>= 140/90) for all patients.[15] Therefore, we used the single threshold of >= 140/90 mm/Hg for all patients in our sample. TI for each subject was calculated using BP values from visits between their respective dates of enrollment until December 2007.

Statistical Analyses. We first assessed distributions for each variable by racial group, performing univariate analyses (t-tests and chi-square, as appropriate). Next, we investigated the effect of race upon TI with and without controlling for patient-level covariates. In these analyses, we used random effects analyses to account for clustering of patients-within-providers. Our multivariate model also included patient race, age, gender, education level, income, number of BP medications and comorbid conditions. For the final model, we added the variables for health beliefs and illness perceptions, perceptions of providers and experiences of discrimination, medication adherence, sodium intake, and provider counseling, keeping only variables that had been significant at the $p \leq 0.10$ level in Model 2. We also performed these regressions without eliminating any variables and found similar results.

To assess the effects of race and TI score on BP, we performed two linear regressions, using systolic BP (SBP) (at the final clinic visit) as the dependent variable. Because uncontrolled BP is mostly a problem of poorly controlled SBP, we used this as our outcome (supplemental analyses found that there were only 166 (1.3%) visits in which patients had a SBP greater than or equal to 140 and diastolic BP less than or equal to 60).[24] The first model included only patient race and the second model also included TI score. All analyses were conducted using SAS 9.1.3 (SAS Institute, Cary, NC).

RESULTS

Black patients were significantly younger, more likely to be female, less educated, had a lower income, prescribed more BP medications compared to white patients, and were more likely to have diabetes, chronic kidney disease, congestive heart failure, and obesity (Table 1). Black patients had significantly more concerns about their BP medication (mean score 2.5 vs. 2.1, p<0.001; Table 2), and believed their BP was more serious, given their current use of medication (mean scores 2.8 vs. 3.3, p<0.001; lower scores indicate greater seriousness). More black patients believed that taking their BP medication would help them to feel better (p<0.001), but fewer of them believed that it would help them live longer, compared to whites (p=0.001).

In the bivariate analyses, important racial differences were noted with regard to almost every hypertension-related belief that we examined, with blacks having less accurate or more negative perceptions (Table 2). For example, black patients were more likely to report that a germ or virus, chance, other people, and poor medical care in the past contributed to causing their high BP. While all patients generally agreed that their provider understood their background and values, black patients agreed less strongly (p=0.03). Also, while all patients disagreed that their provider looks down on them and the way they live their life, white patients disagreed more strongly (p< 0.001). Blacks were more likely to have missing adherence data, generally due to failure to return MEMS caps. They were also more likely to have fair or poor adherence and less likely to have excellent adherence.

Table 1. Sociodemographic and Clinical Characteristics

Patient Characteristics	All patients (n=819)	Black (n=476)	White (n=343)	p-value
Mean age	59.6	58.3	61.4	<0.001
Gender (% male)	33.9	26.7	44.0	<0.001
Education (% less than 12th grade completed)	54.2	65.1	39.2	<0.001
Income (% less than $20,000/year)	48.4	57.4	36.1	<0.001
Mean number of blood pressure medications	2.3	2.4	2.3	0.003
Benign Prostatic Hypertrophy	3.7	1.7	6.4	<0.001
Cerebrovascular Disease	5.6	4.4	7.3	0.08
Chronic Kidney Disease	6.7	8.4	4.4	0.02
Congestive Heart Failure	3.5	5.0	1.5	0.006
Coronary Artery Disease	12.8	9.2	17.8	<0.001
Diabetes	33.2	39.1	25.1	<0.001
Hyperlipidemia	53.6	48.3	60.9	<0.001
Nicotine Dependence	7.5	8.6	5.8	0.14
Obesity	58.9	63.9	52.2	<0.001
Peripheral Vascular Disease	5.3	3.9	7.0	0.057

Table 2. Patient Beliefs and Experience Variables by Race

Patient Variables	Total	Black	White	p-value
Health Beliefs & Illness Perceptions				
Concerns about BP Medication (mean)[a]	2.3	2.5	2.1	<0.001
Beliefs about necessity of BP Medication (mean) [b]	3.7	3.7	3.7	0.15
Patient Beliefs about Blood Pressure (mean) [c]				
How serious is your high BP, in general?	1.5	1.5	1.5	0.28
How serious is your high BP, given your current use of medication?	3.0	2.8	3.3	<0.001
If no BP meds over the next year, would BP get worse?	1.5	1.5	1.4	0.18
If no BP meds over the next year, would develop other health problems?	1.7	1.7	1.7	0.27
Do you believe that taking BP medications will...(% yes)				
make you feel worse?	2.1	2.3	1.8	0.57
help you feel better?	93.7	96.2	90.3	<0.001
help you live longer?	93.8	91.3	97.1	0.001
improve the quality of your life?	93.1	92.6	93.8	0.5
prevent future high BP related illnesses?	93.6	92.9	94.6	0.33
Illness Perception Questionnaire [d]				
Germ or virus caused my high BP	3.9	3.7	4.2	<0.001
Diet played major role in causing my high BP	2.2	2.1	2.4	0.005
Pollution of environment caused my high BP	3.7	3.6	3.8	<0.001
My high BP is hereditary	2.0	1.9	2.1	0.01
It was just by chance that I became ill with high BP	3.5	3.3	3.7	<0.001
Stress was major factor in causing my high BP	2.4	2.4	2.4	0.97
My high BP is largely due to my own behavior	2.7	2.8	2.7	0.27
Other people played large role in causing my high BP	3.4	3.3	3.5	0.005
My high BP caused by poor medical care in the past	3.9	3.7	4.1	<0.001
My state of mind played a major part in causing my high BP	3.2	3.2	3.2	0.82
Perceptions of Provider (mean scores) [d]				
Provider treats me with respect and dignity	1.3	1.3	1.3	0.1
Provider understands my background and values	1.5	1.5	1.4	0.03
Often feel provider looks down on me and the way I live my life	4.4	4.3	4.5	<0.001
Provider understands my cultural background and how it affects my health	1.8	1.8	1.8	0.38
Experiences of Discrimination (% reported any discrimination)	19.9	28.5	7.9	<0.001
Sodium intake score (mean) [e]	5.5	5.5	5.5	0.96
Medication Adherence (%)[f]				
missing	18.1	23.5	10.5	<0.001
poor	6.9	8.2	5.3	
fair	12.8	15.1	9.6	
excellent	62.2	53.2	74.6	
Provider Counseling (mean)[g]	6.6	7.2	5.7	<0.001
Treatment Intensity score (mean)[h]	−0.28	−0.31	−0.24	<0.001
Baseline Systolic BP (mean)	133.6	135.1	131.4	0.002
Final Systolic BP (mean)	132.9	134.2	131.0	0.009

[a] *High score indicates more concerns; range from 1–5*
[b] *High score indicates greater beliefs in necessity of medications; range from 1–5*
[c] *High score indicates less serious or less likely; range from 1–5*
[d] *Higher score indicates more disagreement with statement; range 1–5*
[e] *Higher score indicates worse diet (more sodium, fast food); range 1–12*
[f] *Poor adherence: <50%, fair: 50–80%, excellent: >80%, measured by MEMS caps for 90-day use*
[g] *Higher score indicates more discussion; range from 0–12*
[h] *Higher score indicates more treatment intensity*

Black patients reported more provider counseling about BP (7.2 vs. 5.7, p<0.001), and were much more likely to report at least one experience of discrimination in the health care setting than whites (29% vs. 8%, p<0.001). Finally, black patients had significantly higher baseline and final systolic BP compared to whites (135.1 mm/Hg vs. 131.4 mm/Hg, p<0.01; 134.2 mm/Hg vs. 131.0 mm/Hg, p=0.009, respectively).

In unadjusted analyses, black patients had less TI than whites, equivalent to approximately one fewer therapy increase per 14 clinic visits (−0.31 vs.−0.24, p<0.001) (Table 2). After including patient sociodemographic variables and clinical characteristics in our regression model, black patients had significantly lower TI, equivalent to approximately one fewer therapy increase per 17 clinic visits (Model 2, β=−0.06, p=0.01; Table 3). In the final model, after adding patient beliefs, perceptions of provider, experiences of discrimination, sodium intake, provider counseling and medication adherence, race was no longer a significant predictor of TI (Model 3, β=−0.02, p=0.5).

This final model revealed several determinants of TI. Patients with hyperlipidemia had increased TI, equivalent to approximately one more therapy increase per 13 clinic visits (β=0.08, p<0.001). Increased patient concerns about BP medications and more provider counseling were each related to lower TI, equivalent to approximately one fewer therapy increase per 17 clinic visits for each unit increase on the concerns scale and one less therapy increase per 100 visits for each unit increase on the provider counseling scale (β=−0.06, p=0.02 and β =−0.01, p=0.001, respectively). Two marginally significant findings indicated that more BP medications were associated with reduced TI (β=−0.02, p=0.054) and more disagreement with the item about

Table 3. Factors Associated with Treatment Intensification

Patient Variables	Model 1		Model 2		Model 3	
	Parameter estimate	p-value	Parameter estimate	p-value	Parameter estimate	p-value
Sociodemographic & Clinical Characteristics						
Black Race	−0.08	<0.001	−0.06	0.01	−0.02	0.51
Male	–	–	−0.006	0.81	–	–
Age	–	–	<−0.001	0.98	–	–
Education	–	–	0.02	0.39	–	–
Income	–	–	−0.001	0.96	–	–
Benign prostatic hypertrophy	–	–	0.02	0.70	–	–
Cerebrovascular disease	–	–	−0.08	0.09	−0.07	0.19
Chronic kidney disease	–	–	0.06	0.23	–	–
Congestive heart failure	–	–	0.09	0.18	–	–
Coronary artery disease	–	–	−0.02	0.65	–	–
Diabetes mellitus	–	–	0.02	0.51	–	–
Hyperlipidemia	–	–	0.07	0.002	0.08	<0.001
Nicotine dependence	–	–	0.01	0.75	–	–
Obesity	–	–	−0.03	0.14	–	–
Peripheral vascular disease	–	–	0.03	0.60	–	–
Number of BP medications	–	–	−0.04	<.001	−0.02	0.054
Health Beliefs & Illness Perceptions						
Concerns about BP Medication	–	–	–	–	−0.06	0.02
Beliefs about necessity of BP Medication	–	–	–	–	0.02	0.41
Patient Beliefs about Blood Pressure						
How serious is high BP, in general?	–	–	–	–	−0.01	0.68
How serious is your high BP, given your current use of medication?	–	–	–	–	0.02	0.09
If no BP meds over the next year, would BP get worse?	–	–	–	–	0.02	0.33
If no BP meds over the next year, would develop other health problems?	–	–	–	–	−0.01	0.42
Do you believe that taking BP medications will...						
make you feel worse?	–	–	–	–	0.14	0.10
help you feel better?	–	–	–	–	−0.03	0.51
help you live longer?	–	–	–	–	0.07	0.19
improve the quality of your life?	–	–	–	–	0.01	0.78
prevent future high BP related illnesses?	–	–	–	–	−0.07	0.17
Illness Perception Questionnaire						
Germ or virus caused my high BP	–	–	–	–	−0.004	0.75
Diet played major role in causing my high BP	–	–	–	–	<−0.001	0.98
Pollution of environment caused my high BP	–	–	–	–	−0.005	0.72
My high BP is hereditary	–	–	–	–	−0.007	0.57
It was just by chance that I became ill with high BP	–	–	–	–	−0.002	0.84
Stress was major factor in causing my high BP	–	–	–	–	−0.001	0.97
My high BP is largely due to my own behavior	–	–	–	–	<0.001	0.95
Other people played large role in causing my high BP	–	–	–	–	0.003	0.79
My high BP caused by poor medical care in the past	–	–	–	–	−0.008	0.56
My state of mind played a major part in causing my high BP	–	–	–	–	−0.001	0.94
Perceptions of Provider						
Provider treats me with respect and dignity	–	–	–	–	0.04	0.18
Provider understands my background and values	–	–	–	–	0.02	0.34
Often feel provider looks down on me and the way I live my life	–	–	–	–	0.04	0.05
Provider understands my cultural background and how it affects my health	–	–	–	–	−0.002	0.91
Experiences of Discrimination	–	–	–	–	0.004	0.91
Sodium intake score	–	–	–	–	−0.003	0.66
Medication Adherence						
Missing	–	–	–	–	−0.03	0.34
Poor	–	–	–	–	.01	0.77
Fair	–	–	–	–	−0.06	0.10
Excellent	–	–	–	–	ref	ref
Provider Counseling	–	–	–	–	−0.01	0.001

feeling one's provider looks down on them was associated with increased TI (β=0.04, p=0.05).

To assess the effects of race and TI on BP, we performed two linear regressions, using the final systolic BP as the outcome. When race was examined alone as a predictor, black patients had a higher final systolic BP (134 mm/Hg for blacks vs. 131 mm/Hg for whites, p=0.009; Table 2). In a model containing race and TI as independent variables, TI was a significant predictor of final BP (final SBP 2.0 mm/Hg lower for each additional therapy increase per 10 visits, p<0.001; Table 4), but race was no longer significant (blacks 1.6 mm/Hg higher than whites, p=0.17).

Table 4. Factors Associated with Systolic BP

	Model 1		Model 2	
	Parameter estimate	p-value	Parameter estimate	p-value
Black race	3.15	0.01	1.59	0.17
Treatment Intensity score	–	–	−2.03	<0.001

DISCUSSION

In this effort to understand racial disparities in BP control, we found that black patients had significantly lower rates of TI compared to whites, even after accounting for differences in their clinical and sociodemographic status. However, racial disparities in TI were explained by the inclusion of patient characteristics, health beliefs, and provider counseling. This finding differs from prior work finding that African Americans were treated more intensively for their hypertension,[14,25] but which did not account for patient health beliefs or interactions with providers. One study defined TI as the number of classes of BP medications,[14] a measure which is likely to be confounded by disease severity.[15]

In our study, having hyperlipidemia was the only significant predictor of increased TI. Patient concerns about BP medications and more provider counseling were associated with reduced TI, suggesting that providers may hesitate to intensify treatment when patients express concerns, or that providers may substitute counseling for TI. This conclusion is consistent with other findings indicating that more discussion between providers and patients about medication issues was associated with a lower likelihood of changing treatment.[26] Patient concerns may be related to being on an increased number of BP medications or lack of trust in providers whom patients may feel look down on them. The racial differences in BP control in our sample suggest that substituting counseling for TI is not an effective strategy to minimize disparities in BP control.

Black patients had a higher systolic BP in unadjusted analyses, but this effect was much attenuated and no longer significant after controlling for TI, suggesting that increasing TI may help to resolve disparities in uncontrolled BP. This notion is partly consistent with prior findings that TI was associated with increased odds of having controlled BP; however, in that study, no interactions were found between race and TI.[13] Our results add to this literature by revealing determinants of TI by race and demonstrating that patient concerns and beliefs about BP and provider counseling are associated with differential rates of TI, and thus are a target for intervention, along with increasing provider awareness about intensifying hypertension therapy.

Our findings should be interpreted within the limitations of our study. As part of the parent study, a subset of providers received an educational intervention, which may have affected provider-patient communication and, in turn, patient beliefs about BP medication. However, our analyses adjusted for nesting of patients-within-providers and thus controlled for differential practice styles that may have been associated with providers' exposure to the intervention. In addition, separate analyses indicated no significant effect of the intervention on counseling or BP (not shown). Our sample is comprised of only white and black patients receiving care at a single urban, safety-net hospital and our findings may therefore not be generalizable to other populations. We do not have information about provider attitudes and beliefs, which have also been shown to also be a key factor in clinical inertia.[26,27] Again, our analysis may account for some of this variability, but not all.

Disparities in hypertension control can be minimized by identifying and addressing modifiable factors that contribute to these differential rates in health outcomes. This study found that patient concerns about antihypertensive medications play a significant role in reducing necessary TI. These results provide some support for a model to explain disparities in BP outcomes. In this model, race contributes to racial differences in beliefs and experiences, which contributes to racial differences in TI, which contributes to racial differences in BP control. Future qualitative research to determine the causes of concerns about BP medications and how those concerns impact TI may be helpful.[28] Such findings could be incorporated into a patient or provider intervention, to help address patient concerns about BP medication. In the ongoing struggle to diminish racial disparities in health outcomes, this study offers insight to potential targets for interventions. Improved patient–provider communication and patient health education may have the potential to reduce racial disparities in TI and ultimately, BP control.

Acknowledgements: *The authors thank Mark Glickman, PhD for his advice regarding statistical analyses.*

Funding sources: *This work was supported by the NIH/National Heart, Lung and Blood Institute grant R01 HL072814 (N. Kressin, PI). Dr. Kressin is supported in part by a Research Career Scientist Award from the Department of Veterans Affairs Health Services Research and Development Service. Dr. Rose is funded by a career development award from the Department of Veterans Affairs Health Services Research and Development Service.*

The views expressed in this article are those of the authors and do not necessarily represent the official views and policies of the Department of Veterans Affairs.

Conflict of Interest: *None disclosed.*

Financial Disclosure: *No authors have identified any financial conflicts of interest.*

Corresponding Author: *Meredith Manze, MPH; VA Boston Healthcare System, Boston, MA, USA (e-mail: meredith.manze@bmc.org).*

REFERENCES

1. **Hertz R, Unger A, Cornell J, Saunders E.** Racial disparities in hypertension prevalence, awareness, and management. Arch Intern Med. 2005;165:2098–104.
2. **Smedley B, Stith A, Nelson A.** Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care. Washington, DC: National Academy Press; 2002.
3. **Cutler JA, Sorlie PD, Wolz M, Thom T, Fields LE, Roccella EJ.** Trends in hypertension prevalence, awareness, treatment, and control rates in United States adults between 1988–1994 and 1999–2004. Hypertension. 2008;52:818–27.
4. **Nesbitt SD.** Hypertension in black patients: special issues and considerations. Curr Hypertens Rep. 2005;7:244–8.
5. **Hajjar I, Kotchen T.** Trends in prevalence, awareness, treatment, and control of hypertension in the United States, 1988–2000. J Am Med Assoc. 2003;290:199–206.
6. **Chobanian A, Bakris G, Black H, et al.** The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treat-

ment of High Blood Pressure: The JNC 7 Report. J Am Med Assoc. 2003;289:2560–72.
7. **Wong M, Shapiro M, Boscardin W, Ettner S.** Contributions of major diseases to disparities in mortality. N Engl J Med. 2002;347:1585–92.
8. **Phillips LS, Branch WT, Cook CB, et al.** Clinical inertia. Ann Intern Med. 2001;135:825–34.
9. **Berlowitz D, Ash AA, Hickey EC, et al.** Inadequate management of blood pressure in a hypertensive population. N Engl J Med. 1998;339:1957–63.
10. **Rose AJ, Shimada SL, Rothendler JA, et al.** The accuracy of clinician perceptions of "usual" blood pressure control. J Gen Intern Med. 2007.
11. **Selby JV, Uratsu CS, Fireman B, et al.** Treatment intensification and risk factor control: toward more clinically relevant quality measures. Med Care. 2009;47:395–402.
12. **Okonofua EC, Simpson KN, Jesri A, Rehman SU, Durkalski VL, Egan BM.** Therapeutic inertia is an impediment to achieving the Healthy People 2010 blood pressure control goals. Hypertension. 2006;47:345–51.
13. **Hicks LS, Fairchild DG, Horng MS, Orav EJ, Bates DW, Ayanian JZ.** Determinants of JNC VI guideline adherence, intensity of drug therapy, and blood pressure control by race and ethnicity. Hypertension. 2004;44:429–34.
14. **Safford MM, Halanych JH, Lewis CE, Levine D, Houser S, Howard G.** Understanding racial disparities in hypertension control: intensity of hypertension medication treatment in the REGARDS study. Ethn Dis. 2007;17:421–6.
15. **Rose A, Berlowitz D, Manze M, Orner M, Kressin N.** Comparing methods of measuring treatment intensification in hypertension care. Circulation Cardiovascular Quality and Outcomes. 2009;2:385–91.
16. **Horne R, Weinman J, Hankins M.** The beliefs about medicines questionnaire: the development and evaluation of a new method for assessing the cognitive representation of medication. Psychol Health. 1999;14:1–24.
17. **Horne R, Buick D, Fischer M, Leake H, Cooper V, Weinman J.** Doubts about the necessity and concerns about adverse effects: identifying the types of beliefs that are associated with non-adherence to HAART. Int J STD AIDS. 2004;15:38–44.
18. **Kressin N, Wang F, Long J, et al.** Hypertensive patients' health beliefs, process of care, and medication adherence: is race important? J Gen Intern Med. 2007;22:768–74.
19. **Weinman J, Petrie KJ, Moss-Mossis R, Horne R.** The illness perception questionnaire: a new method for assessing the cognitive representation of illness. Psychol Health. 1996;11:431–45.
20. **Johnson RL, Saha S, Arbelaez JJ, Beach MC, Cooper LA.** Racial and ethnic differences in patient perceptions of bias and cultural competence in health care. J Gen Intern Med. 2004;19:101–10.
21. **Bird ST, Bogart LM.** Perceived race-based and socioeconomic status (SES)-based discrimination in interactions with health care providers. Ethn Dis. 2001;11:554–63.
22. **Kim MT, Hill MN, Bone LR, Levine DM.** Development and testing of the hill-bone compliance to high blood pressure therapy scale. Prog Cardiovasc Nurs. 2000;15:90–6.
23. **Pbert L, Adams A, Quirk M, Hebert J, Ockene J, Luippold R.** The patient exit interview as an assessment of physician-delivered smoking intervention: a validation study. Health Psychol. 1999;18:183–8.
24. **Hyman DJ, Pavlik VN.** Characteristics of patients with uncontrolled hypertension in the United States. N Engl J Med. 2001;345:479–86.
25. **Umscheid CA, Gross R, Weiner MG, Hollenbeak CS, Tang SS, Turner BJ.** Racial disparities in hypertension control, but not treatment intensification. Am J Hypertens. 2010;23:54–61.
26. **Kerr EA, Zikmund-Fisher BJ, Klamerus ML, Subramanian U, Hogan MM, Hofer TP.** The role of clinical uncertainty in treatment decisions for diabetic patients with uncontrolled blood pressure. Ann Intern Med. 2008;148:717–27.
27. **Ferrari P.** Reasons for therapeutic inertia when managing hypertension in clinical practice in non-Western countries. J Hum Hypertens. 2009;23:151–9.
28. **Bokhour B, Long J, Berlowitz D, Kressin N.** Assessing patient adherence in medical encounters: How do providers talk with patients about antihypertensive medication taking? J Gen Intern Med. 2006;21:577.

Original Article

The Path to Physician Leadership in Community Health Centers: Implications for Training

Jeffrey F. Markuns, MD, EdM; Bruce Fraser, PhD; Jay D. Orlander, MD, MPH

Background and Objectives: *Community health centers are facing a shortage of primary care physicians at a time when government plans have called for an expansion of community health center programs. To succeed with this expansion, community health centers require additional well-trained physician leadership. Our objective was to ascertain how medical directors obtain leadership skills in an attempt to identify the best methods and venues for providing future leadership training programs.* Methods: *Using recorded interviews and focus group data with community health center medical directors, we identified patterns and themes through cross-case content analysis to determine leadership training needs in underserved settings.* Results: *Medical directors often enter positions unprepared and can quickly become frustrated by an inability to make system improvements. Medical directors seek multiple ways to obtain the leadership skills necessary, including conferences, peer networking, mentorship, and formal degree training. Many directors express a desire for additional training, preferring flexibility in curriculum and hands-on components.* Conclusions: *Additional leadership training opportunities for active and future medical directors are needed. Academic medical centers and other training sponsors should consider innovative ways to develop effective physician leadership to provide quality care to underserved communities.*

(Fam Med 2010;42(6):403-7.)

Community health centers (CHCs) take a lead role in providing quality care to the underserved. The Health Resources and Services Administration (HRSA) Bureau of Primary Care's Community Health Center Program provides care for more than 16 million underserved people and has undergone rapid expansion.[1-3] CHCs provide high-quality cost-effective care to underserved communities, improve access to care, and reduce health care disparities.[4]

For CHCs to continue to expand in the face of limited resources, strong physician leadership is needed. However, there may not be a sufficient qualified workforce.[5-8] Many active CHC medical directors may also lack sufficient training in practice management and the specific skills such leadership requires.[9-14]

Although medical directors play an important role in direct patient care and CHC management, there is limited research elucidating their basic characteristics, such as roles and responsibilities, relationships with other administrators, retention, and satisfaction levels.[10,11,15] Some studies suggest administrative inexperience and insufficient training may lead to poor retention of medical directors.[10,11] Further, increased training correlates with improved leadership skills, as reported by CHC executive directors,[13] and leadership skills play a significant role in achieving exemplary CHC practices.[16] Guidance on how medical directors should obtain such leadership training, however, is lacking.

Our primary objective was to ascertain the need for leadership training programs for medical directors of CHCs. This included determining how medical directors obtain skills in practice management and leadership as well as the most appropriate methods and venues for this training.

From the Department of Family Medicine (Dr Markuns), School of Education (Dr Fraser), and Evans Department of Medicine (Dr Orlander), Boston University; and VA Boston HealthCare System, West Roxbury, MA (Dr Orlander).

Methods

We performed a qualitative study analyzing key informant interviews with current and former CHC medical directors. Semi-structured interviews, audio recorded and later transcribed, were performed over a period of 12 months during 2006–2007 by the same interviewer using a standardized question guide. The validity of the insights gained from these interviews were then evaluated through a focus group of a separate sample of medical directors. The study was approved by the Boston University School of Medicine Institutional Review Board.

Samples

We recruited a purposeful sample of 11 current and former medical directors through a hospital-affiliated urban network of independent CHCs in Boston. We specifically chose participants based on current and previous experience and community demographics to achieve a diverse group. All medical directors approached for interviews agreed to participate except one.

For the focus group, four current and former medical directors were recruited from a training program sponsored by the Massachusetts League of Community Health Centers. All were affiliated with CHCs outside of the Boston-based hospital-affiliated network that served as the source of key informants.

Instruments

We developed an initial question guide based upon literature review, personal experience of the authors, and discussion with individuals who provide medical director training, including leaders from state and national CHC organizations. After the first three recorded interviews with medical directors, the question guide was further refined to better address the core objectives of the study.

Analysis

All transcripts were coded and analyzed for common themes and higher-order categories using the qualitative research software NVivo 7 to perform cross-case content analysis. We performed pattern and theme recognition based on inductive and deductive processes using grounded theory principles in an iterative process and performed interviews until data saturation was achieved.[17] Coded data was reviewed by the coauthors, neither of whom is affiliated with CHCs. Focus group responses were analyzed in a similar manner.

Results

Participant Demographics

Of the 15 medical directors participating, 12 were active medical directors at the time of the interviews. The remaining medical directors were previous medical directors now involved in administrative roles outside of CHCs; however, all continued to work closely with CHC medical directors. Ten of the medical directors were men. Although a majority were Caucasian, the sample did include physicians of multiple racial and ethnic backgrounds, not specifically identified so that we could maintain confidentiality.

Medical directors interviewed had a range of experience in the position, ranging from 1 month to 35 years (median 3 years). Time since graduation from residency ranged from 5 to 35 years (median 14.5 years). All were primary care physicians: nine were family physicians, three were pediatricians, three were internists, and one was trained in both internal medicine and pediatrics.

Content Analysis

Three prominent themes emerged in relation to leadership training: (1) motivation for improvement as the path to leadership, (2) inexperience as a barrier to achieving change, and (3) training as a facilitator for success.

Theme 1: Path to Leadership: Motivation for Improvement

Our subjects typically followed a path to leadership driven by a desire for health care improvement. Although some physicians were recruited to the medical director position immediately upon completing postgraduate training, most were not. More often, directors were recruited to leadership positions after becoming established in an initial clinical role either in a CHC or elsewhere. Most directors accepted their first clinical position in a CHC or other underserved setting, citing motivators such as a desire to participate in social change or community health. Respondents express a desire to make system improvements in care and as a result evolved into leadership positions. Some felt recruited or pulled into a leadership position by other CHC administrators. One director said:

> I started off just being a pediatrician, but because of my nature of being a busybody and wanting to improve things, I sort of got sucked into the administrative thing. (Director A)

Another stated that:

> I always felt that I would pay back the community for its investment in me, and so I had always planned to be involved in community medicine…A position opened up as medical director, and I thought I would like to try that…I was looking for a way to kind of have some influence beyond individual treatment…so I thought, you know, I might be able to do that being a medical director. (Director G)

Theme 2: Inexperience: A Barrier to Achieving Change

Once established, medical directors almost universally describe feeling unprepared for a poorly defined role. Many report little prior knowledge of CHCs, and only a few report a clear understanding of the role of a medical director. Job descriptions are often absent or considered outdated or inadequate. Even highly experienced medical directors have difficulty providing a focused, concise summary of their role. They cite extremely broad areas of responsibility encompassing multiple areas of leadership, sometimes overlapping the supervisory duties of other CHC administrators. One director reported that:

> I remember telling him, 'Do I get a job description on this?' I mean, I had a sense. He said, 'Ah, we can—it won't help.' He probably would have been right, because I don't think he had a great idea. (Director F)

Another said:

> I don't even know, I don't know what my job is, if I tell you the truth. (Director J)

While some medical directors accepted their position without formal leadership training, others had participated in fellowships or earned additional graduate degrees prior to becoming medical directors. Regardless of prior experiences, almost all frequently felt unprepared in the early phases of their careers as medical directors. The combination of inexperience and lack of direction in this period often acted as a barrier to achieving improvements within their CHCs.

As desire for improvement is a primary motivator, the ability to effect successful change within a CHC appears to play a large role in the satisfaction of medical directors. Medical directors unable to achieve successful CHC improvements express considerable frustration, often including a desire to leave the position, potentially contributing to troubles with retention and turnover. As one director said:

> I took over one of the sites as the local site director and burned out in about a year. Just totally flamed out. Cause I wasn't really ready. My skills weren't really at the level that they needed to be. I couldn't do it. (Director C)

Another reported that:

> This is a place I see right now, as we stand, if we don't make changes, it's a place that you go the last five years before you retire. When you don't give a damn about anything. So that's what's bothering me right now. And that's a big issue. At least for me. Because I need to move on. I need to move on. (Director J)

Theme 3: Training: a Facilitator for Success

To address the barrier of inexperience, we explored the role of subsequent administrative education. Training fell into four main categories: conferences, peer networking, mentorship, and formal degree programs. Most medical directors attended at least one management conference designed specifically for CHC medical directors. One of the focus group participants stated that:

> I can't do this without some training, or I'm just gonna fail right away.

Medical directors frequently referred to peer networking as a potential method for skill enhancement. It is not clear, however, if medical directors use this primarily to discuss mutual areas of inexperience or if they actively benefit from each others' experiences as surrogate knowledge. Peer networking takes place primarily through conferences and electronic mail. A sample comment was that:

> So we can sit and just sort of talk about things in general, and get a sense of where each other are, call each other and ask for help or advice, or to discuss a topic. We have a listserve…where people put out questions to the whole medical director group, and people are very good about responding. To me, that's been very powerful... (Director H)

The guidance of a mentor is often mentioned as a highly effective training method. Those with mentors often praise the relationship. Some participants describe trying to provide mentoring for others, even when they lack such assistance themselves. When available, mentoring appears to be one of the most effective methods for achieving success and satisfaction in the position. Director A stated that:

> Mentors are the only way I have survived.

Director C said that:

> He was just a senior person, and…he was very helpful in terms of commenting on how the world works…I would say, 'This thing just seems funny' and he would say, 'Well, this is why it's not funny. This is why it really is how it is...' And it was really useful to get reality checks from people like him.

Several of the medical directors interviewed pursued graduate degrees. In general, these are felt to be very helpful, particularly those focusing upon business or health care administration. A formal degree seems most helpful when obtained after acquiring some experience as a physician administrator, rather than prior to any such work experience.

> I started working as the site director again a little while after I started business school. And I was much better able to do it after a while…the job just came at me slower. And so that was a much more enjoyable year. (Director C)

When asked about the types of training they would desire most, subjects preferred programs with a practical approach. Medical directors feel programs should allow them to directly apply training to problems in their own health centers. They also seek flexibility in adapting such training to fit their personal needs and commitments. Sample comments included:

> It'd be nice to be able to take that [conference] stuff home, work on it, and get more feedback…later on. (Director D)

> It could be, spend three hours on a Saturday morning… do that once a month, and then have time between courses where you live your life…Folks'll learn this interesting concept, and then carry it back to where they work, and then have enough time to noodle around with it while they're doing their life. (Director C)

In the focus group, core themes from the individual interviews were confirmed. There was a particular focus on preferred training approaches.

In summary, medical directors entered leadership roles motivated by a desire to make improvements in CHC care. Many quickly found themselves unprepared and unable to achieve improvements they desired. Many sought training through conferences, peer networking, mentoring, and formal degree programs. In general, medical directors believe that additional training programs would be useful, especially if programs included longitudinal project work with direct application, components of peer networking and mentorship, and flexibility in interacting with the curriculum.

Discussion

In recent years, the medical community has focused attention on health care quality, patient safety, chronic disease management, operational improvement, and information management. The complexity of these tasks demands a higher level of leadership skill than ever before,[18] yet many new physician leaders of CHCs feel unprepared. Our results show this to be compounded by poorly defined roles, lack of mentorship or peer networking opportunities, and medical directors' own inexperience.

A resulting lack of success by medical directors in achieving CHC improvement may contribute to human resource issues of retention and turnover. To be successful, medical directors must provide transformative leadership to their institutions: training can assist in achieving this goal.[13]

Our findings suggest training methods that may be most useful. Longitudinal programs that provide mentoring and project work are preferred by our participants. A number of conference training opportunities are available, but these 1-day to 1-week courses do not suffice. The Health Disparities Collaboratives developed by HRSA and the Institute for Healthcare Improvement is one model that includes longitudinal project work and provides a number of helpful resources.[19] It has the broad mission of health care system improvement focused on health disparities but does not focus specifically on the needs of CHC medical directors. We believe our findings demonstrate that additional accessible and comprehensive leadership training opportunities are needed specifically for the clinical leaders of these institutions.[20]

Although our data are limited by the urban nature of our sample, we expect these themes remain similar in other settings. While it is possible that physician leaders in rural CHCs have a different set of experiences and preferences, given inherent challenges in rural areas of accessing educational resources, it seems likely that rural medical directors would share many of the same frustrations regarding training and overall leadership preparedness. It also is important to note that our sample is recruited almost entirely from Massachusetts, an area with a high number of tertiary care medical centers and robust funding mechanisms of health care for the uninsured relative to many other areas of the country. These differences would seem to make it more likely, however, that our participating medical directors would have good access to postgraduate leadership training opportunities and fewer financial, recruitment, and other challenges relative to counterparts in other areas of the nation.

This research has important implications for medical education as well as public policy. At a time when comprehensive health care reform is being discussed, the fate of the medically underserved becomes inextricably entwined with the health of the entire nation, and CHC medical directors are a key group of primary care physician leaders necessary for providing high-quality health care to underserved populations. The struggles of these physician leaders may signal a more significant deficit in medical education regarding issues of physician leadership, particularly in primary care. Our results suggest further training is needed both to prepare new medical directors and successfully retain those currently serving in this role. Government, health centers, and academic and postgraduate training institutions will need to work together to generate adequate funding and develop and coordinate this important leadership training to provide the highest quality health care to all.

Acknowledgments: Our thanks to the medical directors and the staff from the Massachusetts League of Community Health Centers who assisted in our research.

This study was presented at the 2008 Society of Teachers of Family Medicine Annual Spring Conference in Baltimore.

This study was financed by HRSA Faculty Development grant D55HP00215.

Corresponding Author: Address correspondence to Dr Markuns, Boston University, Department of Family Medicine, Dowling 5 South, 1 Boston Medical Center Place, Boston, MA 02118. jeffrey.markuns@bmc.org.

References

1. The safety net on the edge. Bethesda, MD: National Association of Community Health Centers, August 2005.
2. Iglehart JK. Spreading the safety net—obstacles to the expansion of community health centers. N Engl J Med 2008;358(13):1321-3.
3. The Health Center Program: The President's Health Center Initiative. http://bphc.hrsa.gov/presidentsinitiative/. Accessed January 10, 2009.
4. Proser M. Deserving the spotlight: health centers provide high-quality and cost-effective care. J Ambul Care Manage 2005;28(4):321-30.
5. Salsberg E, Grover A. Physician workforce shortages: implications and issues for academic health centers and policymakers. Acad Med 2006;81(9):782-7.
6. Green LA, Phillips RL Jr. The family physician workforce: quality, not quantity. Am Fam Physician 2005;71(12):2248, 2253.
7. Rosenblatt RA, Andrilla CH, Curtin T, Hart LG. Shortages of medical personnel at community health centers: implications for planned expansion. JAMA 2006;295(9):1042-9.
8. Freeman J, Ferrer RL, Greiner KA. Viewpoint: developing a physician workforce for America's disadvantaged. Acad Med 2007;82(2):133-8.
9. Halpern R, Lee MY, Boulter PR, Phillips RR. A synthesis of nine major reports on physicians' competencies for the emerging practice environment. Acad Med 2001;76(6):606-15.
10. Shi L, Samuels ME, Cochran CR, Glover S, Singh DA. Physician practice characteristics and satisfaction: a rural-urban comparison of medical directors at US community and migrant health centers. J Rural Health 1998;14(4):346-56.
11. Cochran C, Peltier JW. Retaining medical directors in community health centers. The importance of administrative relationships. J Ambul Care Manage 2003;26(3):250-9.
12. Crites GE, Schuster RJ. A preliminary report of an educational intervention in practice management. BMC Med Educ 2004;4:15.
13. Xirasagar S, Samuels ME, Curtin TF. Management training of physician executives, their leadership style, and care management performance: an empirical study. Am J Manag Care 2006;12(2):101-8.
14. Shepperd JD, Jr. Nontraditional graduate training for administrators of neighborhood health centers. Public Health Rep 1976;91(5):452-7.
15. Samuels ME, Cochran CR, Shi L. A profile of women medical directors in community and migrant health centers. J Ambul Care Manage 2001;24(1):84-91.
16. Craigie FC Jr, Hobbs RF III. Exploring the organizational culture of exemplary community health center practices. Fam Med 2004;36(10):733-8.
17. Walker D, Myrick F. Grounded theory: an exploration of process and procedure. Qual Health Res 2006;16(4):547-59.
18. Schwartz RW, Pogge CR, Gillis SA, Holsinger JW. Programs for the development of physician leaders: a curricular process in its infancy. Acad Med 2000;75(2):133-40.
19. Health Disparities Collaboratives. www.healthdisparities.net/hdc/html/home.aspx. Accessed May 2, 2008.
20. Markuns JF, Culpepper L, Halpin WJ, Jr. Commentary: A need for leadership in primary health care for the underserved: a call to action. Acad Med 2009;84(10):1325-7.

Informatics in Primary Care 2010;18:59–68

Refereed paper

Disparities in health-related internet use by US veterans: results from a national survey

D Keith McInnes ScD MS
Research Health Scientist at the Center for Health Quality, Outcomes and Economic Research (CHQOER), Department of Veterans Affairs (VA) Bedford Medical Center, Bedford, MA, USA and Research Assistant Professor, Department of Health Policy and Management, Boston University School of Public Health, Boston, MA, USA

Allen L Gifford MD
Associate Director, CHQOER, VA Bedford Medical Center, Bedford, MA, USA and Associate Professor, Department of Health Policy and Management, School of Public Health, and Section of General Internal Medicine, School of Medicine, Boston University, Boston, MA, USA

Lewis E Kazis ScD
Senior Scientist, CHQOER, VA Bedford Medical Center, Bedford, MA, USA and Director, Center for the Assessment of Pharmaceutical Practices (Health Outcomes Technologies Program) and Pharmaceutical Assessment Management and Policy Educational Program (PAMP), Department of Health Policy and Management, and Professor, Department of Health Policy and Management, Boston University School of Public Health, Boston, MA, USA

Todd H Wagner PhD
Health Economist, VA Palo Alto Health Care System and Consulting Associate Professor in Health Research and Policy, Stanford University School of Medicine, Palo Alto, CA, USA

ABSTRACT

Background The internet can contribute to improved access to information and services among underserved populations. Little is known about veterans' use of the internet for health, and how it is affected by socio-demographic characteristics. This knowledge gap is acute given the US Department of Veterans Affairs' (VA's) deployment of a major patient portal/personal health record system.
Objectives To assess the frequency and correlates of veterans' use of the internet and identify personal characteristics impeding veterans' health-related internet use.
Methods Survey of 12 878 randomly selected adults from a panel of 60 000 US households. Veterans were oversampled.
Results Of the 3408 veterans responding, 54% had used the internet and 29% had used the internet specifically for health. In multi-variable analyses, general internet use was positively associated with younger age (OR = 0.03, CI = 0.01–0.06, oldest versus youngest group), higher income (OR = 3.12, CI = 2.10–4.63, ≥$75 000 versus <$25 000), more education (OR = 4.2, CI = 2.92–6.02, most versus least educated group), and better health (OR = 0.59, CI = 0.42–83, fair/poor versus very good/excellent). Health-related internet use was positively associated with more education (OR = 2.32, CI = 1.45–3.74, most versus least educated group), urban location (OR = 2.41, CI = 1.66–3.50), and worse health (OR = 1.85, CI = 1.16–2.95, fair/poor versus very good/excellent).
Conclusions In the first large, systematic survey of veterans' internet use we found that more education and urban location were strongly, and positively, associated with veterans' health-related internet use, even after controlling for multiple socio-demographic characteristics. Interventions may be needed for less educated and rural veterans, e.g. by providers discussing internet use with their patients, or by the VA training veterans in health-related internet use.

Keywords: disparities, health information, internet, veterans

Introduction

Access to health information helps patients to make health-related decisions, communicate with their providers and enhance their self-care skills.[1–3] The internet is a promising means of delivering health information to patients[4] and allowing them better access to healthcare systems.[5] Additionally 70–80% of Americans use the internet.[6,7] Studies examining health-related internet use estimate wide differences in use. The Pew Internet and American Life Project has estimated that 63% of Americans have looked for health information online.[6] A recent study, however, looking at use in past 30 days, found that only 13% of Americans had sought health-related information.[8] Additionally, while some disparities seem to consistently emerge, e.g. females and those more educated being more likely to search the internet for health information,[9] there is less consensus about race/ethnicity[10] and very little information about the role of rural location.[11]

Understanding patterns of internet use for health are important for the healthcare organisations that are making the internet an integral component of communication, services and information for patients, i.e. through personal health record (PHR) systems. Two well known examples of PHRs are 'My Health Manager' at Kaiser Permanente and 'MyGroupHealth' at Group Health Cooperative, both of which allow patients to view parts of their electronic medical record and to complete transactions, such as ordering prescription refills or sending secure emails to their doctor or nurse.[5,12] These types of internet-based systems, while they may improve access in general, do not necessarily overcome existing disparities. A study of the Kaiser Georgia PHR found, for example, that African Americans are less likely to register for the Kaiser PHR than white Americans, even after controlling for education, income and internet access.[10]

The Veterans Health Administration (VHA), one of the largest healthcare systems in the USA, also invested substantially in a PHR called My HealtheVet. My HealtheVet provides veterans with health information, allows them to create and maintain health logs, enables online refilling of prescriptions and sends electronic reminders for preventive tests and screening. Despite outreach efforts only about 15% of the approximately five million veterans who receive care in the VHA have registered for My HealtheVet.[13] The reasons are not well understood, and to date there have been no large systematic surveys of veterans' use of the internet for health, though studies of select subgroups of veterans have found that between 54 and 77% have access to the internet.[14,15] We sought to examine the characteristics of veterans and their use of internet, both in general and for health-related purposes. We had access to data from a large national survey conducted in 2002 that oversampled veterans and included questions about health-related internet use.

The survey provides important baseline data for the development of My HealtheVet and similar PHR efforts in other healthcare systems. Rates of general internet use will have assuredly increased since the survey was conducted, but socio-demographic disparities in use of technologies tend to persist for long periods of time.[16] Knowledge of such disparities is crucial for health planners interested in expanding healthcare access through PHRs to hard to reach groups. We were particularly interested in rural–urban disparities in use of the internet and use of the internet for health because serving rural veterans remains a challenge, and a priority, for the VA.[17] We conducted this study first to examine rates of general internet use among veterans, and the socio-demographic characteristics associated with that use; and second to examine, among those veterans who used the internet, the characteristics associated with health-related internet use.

Methods

The study data were collected as part of a larger national survey of internet use for health which reported findings about the general population's use of the internet for health,[18,19] and about internet use among persons with stigmatised illness[20] and among persons with chronic conditions.[21] The study was based on a research panel of 60 000 US households created and maintained by Knowledge Networks Inc., Menlo Park, California. Potential households were contacted using random digit dialling. As part of study participation, every panel household received WebTV internet hardware and software.

For this study, a survey was sent via the internet to 12 878 randomly selected panel members aged 21 years or older, including oversampling of veterans and persons of 50 years of age and older. The survey sample was comparable to the National Health Interview Survey and the Census Bureau's Current Population Survey.[18] Data were collected between December 2001 and January 2002. A total of 8935 persons (69.4%) completed the survey. Additional details of the research panel and sampling have been previously reported.[18]

Respondents were asked about their use of the internet for health-related purposes. All were asked, at the time they joined the panel, whether they had ever used the internet prior to receiving WebTV; those who said yes were classified as 'internet users'. Of the 8935 survey respondents, 3408 were US veterans and formed the analytical sample for this study. Figure 1

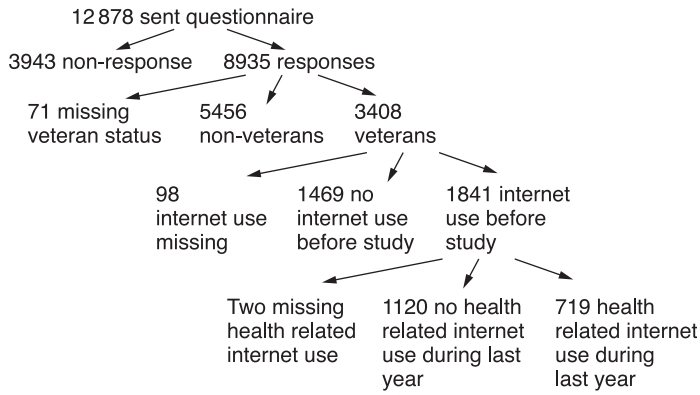


Figure 1 Diagram of analysis of veterans' use of internet for health-related behaviour

shows how the veteran sub-sample is associated with the overall study sample.

Respondents were surveyed about a range of socio-demographic, behavioural and health characteristics, as well as about their use of the internet. Full text of specific items is available from the authors on request. Veterans were identified by a question about being discharged from active duty in the US Army, Navy, Air Force, Marine Corps or Coast Guard. Internet use referred to respondents reporting ever having used the internet prior to the start of the study. Health-related internet use was use of the internet for information or advice about health or health care (possible responses: more than once a week, about once a week, once a month, every two to three months, less than every two to three months or never). For regression analysis we dichotomised this variable with 1 = yes (had used the internet for health in the past year), and 0 = no (had not used the internet for health in the past year).

Race/ethnicity categories were White non-Hispanic, Black non-Hispanic, Hispanic, other non-Hispanic (which included Asian, American Indian or Alaska Native and Pacific Islander). In logistic analyses it was dichotomised (1 = white, 0 = other) due to the small percentages of non-white respondents. Health status was assessed with an item asking, 'Would you say your health in general is ...?', with possible responses of excellent, very good, good, fair or poor, which we recategorised to excellent/very good, good, or fair/poor because of the small percentages in the excellent and poor categories.

Urban location was based on the federal government's categorisation of Metropolitan Statistical Areas, which are urbanised areas of high population density. Travel time to medical care was determined by the item, 'When you need medical care, how long does it take you to get to the place you usually go for care?', with responses of <15 minutes, 15 to 29 minutes, 30 to 60 minutes, and >60 minutes. In logistic analyses we dichotomised this variable (1 = 30 or more minutes, 0 = less than 30 minutes).

Marital status, originally given five categories, was dichotomised (1 = married, 0 = single/divorced/widowed/separated). Number of chronic conditions refers to diagnoses of hypertension, diabetes, cancer, depression and heart problems. We created a dichotomous variable for these (1 = one or more; 0 = none). Use of a VHA hospital or clinic was determined by an item that asked 'Have you received any medical care at a VA hospital or clinic in the last two years?' (1 = yes, 0 = no).

We compared veterans to non-veterans on a number of demographic, health services and internet-related characteristics using chi-square analysis. Among veterans we compared health-related internet use between users and non-users of VHA services. We used multivariable logistic regression to examine, among all the veterans in the study, the characteristics associated with internet use. Next, among the subset of veterans who were internet users, we examined the characteristics associated with health-related internet use. We reported adjusted odds ratios and 95% confidence intervals. In all analyses we used post-stratification weights to match the respondents to the known distribution of the US population on age, sex, race, education, region, metropolitan residence and veteran status, and to account for the oversamples and for non-response. We corrected the standard errors for the complex survey design using PROC SURVEYFREQ and PROC SURVEYLOGISTIC procedures in SAS 9.1 (SAS Institute, Inc., Cary, NC). Chi-square estimates were based on Rao–Scott chi-square tests which adjust for the complex survey design. The study was approved by the Institutional Review Boards at Stanford University and at the Edith Nourse Rogers Memorial VA Hospital, Bedford, MA.

Results

Table 1 compares the demographic and health-related characteristics of veterans and non-veterans. Veterans and non-veterans were equally likely to have used the internet (53.9 vs. 52.5%, $P = 0.41$). Veterans were older, and more often white, married and male ($P<0.001$ for each). They had higher incomes and more education ($P<0.001$ for each). Although more likely to have a chronic health condition, veterans reported similar overall health status to non-veterans.

Nearly one-third of all respondents nationally reported using the internet to search for health information, with veterans reporting similar health-related internet use (29.2%) to non-veterans (32.5%, $P = 0.18$) (Table 2). About 7–8% of both veterans and non-veterans used the internet frequently (monthly or more often) for health-related information. Among veterans, both users and non-users of VHA care were equally likely to use the internet for health-related information ($P = 0.63$).

To understand the factors associated with veterans' internet use, we constructed multivariate regression models to identify independent predictors of internet use in general, and of internet use for health in particular (Table 3). Age was important for general internet use, with the oldest veterans significantly less

Table 1 Characteristics of respondents to the Health Care and Internet survey ($n = 8864$),[a] weighted frequencies

	Non-veterans ($n = 5456$)	Veterans ($n = 3408$)	Chi-sq.[b]	*P*
Age, yrs ($n = 8864$)			404.6	<0.0001
21–34	26.0%	8.9%		
35–49	37.2%	22.2%		
50–64	23.1%	35.7%		
65–74	9.2%	21.0%		
75+	4.6%	12.3%		
Race/ethnicity ($n = 8864$)			81.2	<0.0001
White	71.7%	82.9%		
Black	11.6%	9.4%		
Hispanic	13.3%	4.9%		
Other	3.4%	2.8%		
Sex ($n = 8864$)			697.3	<0.0001
Male	38.6%	93.9%		
Female	61.4%	6.1%		
Household income, $ ($n = 8015$)			69.7	<0.0001
<25 000	29.9%	17.3%		
25–49 000	37.3%	40.6%		
50–74 000	19.6%	24.2%		
75 000+	13.2%	17.9%		
Education, yrs ($n = 8864$)			106.1	<0.0001
<13	60.1%	45.1%		
13–16	33.8%	46.8%		
>16	6.2%	8.1%		
Urban vs non-urban ($n = 8864$)			0.07	0.80
Urban	77.7%	77.2%		
Non-urban	22.4%	22.8%		
Travel time to medical care ($n = 8802$)			0.16	0.92
<15 min	49.4%	48.7%		
15–29 min	36.7%	37.2%		
≥30 min	14.0%	14.1%		

Table 1 Continued

Marital status (*n* = 8788)			51.7	<0.0001
Married	61.4%	72.6%		
Single	18.0%	8.6%		
Divorced	12.0%	12.0%		
Widowed	6.4%	5.0%		
Separated	2.2%	1.9%		
Health status (*n* = 8850)			2.5	0.28
Fair/poor	16.1%	16.6%		
Good	36.9%	34.4%		
Very good/excellent	47.0%	49.0%		
Number of chronic conditions (*n* = 8864)			40.4	<0.0001
Chronic conditions 1+	48.8%	59.4%		
Chronic conditions 0	51.2%	40.6%		
Used VA hospital/clinic in last 2 yrs (*n* = 8864)			1307.0	<0.0001
Yes	0.0%	16.8%		
No	100.0%	83.2%		
Used internet (prior to WebTV) (*n* = 8583)			0.68	0.41
Yes	52.5%	53.9%		
No	47.5%	46.1%		

[a] *n* for some variables not equal to 8864 due to item missing values
[b] Rao–Scott Chi-square for all comparisons except for 'Used VA hospital/clinic in last 2 yrs' for which it was not calculable due to a cell with zero observations. Instead Pearson Chi-square was estimated.
yrs = year
min = minutes
VA = Department of Veteran Affairs

Table 2 Frequency of use of internet for health information, by veteran status, weighted

	Monthly or more often (%)	Every 2–3 mos (%)	<Every 2–3 mos (%)	Never (%)	Chi-sq[a]	*P*
General population (*n* = 8859)[b]					5.49	0.18
Non-vets (86.0%)	7.5	5.0	20.0	67.6		
Veterans (14.0%)	7.3	4.4	17.5	70.7		
Among veterans (*n* = 3406)[c]					1.74	0.63
VHA user (16.8%)	6.6	4.6	15.3	73.4		
Non-VHA user (83.8%)	7.5	4.4	18.0	70.2		

[a] Rao–Scott Chi-sq, df = 3
[b] *n* does not equal 8864 due to missing values for 'internet use for health'
[c] *n* not equal to 3408 due to missing values for 'internet use for health'
mos = months
VHA = Veterans Health Administration

Table 3 Veteran use of internet and use of internet for health-adjusted odds ratio

	Use of internet among veterans ($n = 3310$)[a]			Use of internet for health among veterans who use internet ($n = 1839$)[b]		
	Adj. OR	CI	*P* value	Adj. OR	CI	*P* value
Age (years)						
21–34						
35–49	0.20	0.10–0.43	<0.001	1.29	0.69–2.42	0.43
50–64	0.10	0.05–0.21	<0.001	1.16	0.67–1.968	0.60
65–74	0.05	0.02–0.10	<0.001	1.31	0.72–2.40	0.38
75+	0.03	0.01–0.06	<0.001	1.00	0.51–1.98	1.00
Race/ethnicity						
White	0.89	0.64–1.22	0.46	0.92	0.61–1.38	0.67
Other						
Sex						
Male						
Female	1.16	0.60–2.24	0.66	1.36	0.71–2.61	0.36
Household income $						
<25 000						
25–49 000	1.26	0.93–1.70	0.13	1.15	0.70–1.88	0.59
50–74 000	1.74	1.24–2.43	0.001	0.84	0.52–1.37	0.49
75 000+	3.12	2.10–4.63	<0.001	1.22	0.72–2.05	0.46
Education years						
<13						
13–16	2.67	2.11–3.37	<0.001	1.87	1.27–2.73	0.001
>16	4.20	2.92–6.02	<0.001	2.32	1.45–3.74	<0.001
Urban vs non-urban						
Urban	0.87	0.64–1.19	0.38	2.41	1.66–3.50	<0.001
Non-urban						
Travel time to medical care						
<30 min.						
30+ min.	1.06	0.77–1.46	0.73	0.89	0.58–1.36	0.58
Marital status						
Married	1.07	0.80–1.43	0.63	0.89	0.62–1.26	0.50
Single[c]						
Health status						
Fair/poor	0.59	0.42–0.83	0.002	1.85	1.16–2.95	0.01
Good	0.69	0.54–0.88	0.002	1.54	1.09–2.17	0.01
Very good/excellent						

Table 3 Continued

Number of chronic conditions						
Chronic conditions 1+	1.22	0.95–1.57	0.12	1.23	0.88–1.71	0.22
Chronic conditions 0						
Used VA hospital/clinic in last 2 years						
Yes	0.87	0.62–1.21	0.40	0.73	0.49–1.07	0.11
No						

[a] Not equal to 3408 due to missing internet use
[b] Not equal to 1841 due to missing health-related internet use
[c] Single refers to never married, divorced, widowed or separated
Urban = resident of metropolitan statistical area (MSA)
Adj. = adjusted OR = odds ratio
CI = confidence interval
VA = Department of Veteran Affairs
Frequencies weighted to correct the distribution of respondents to match the known distribution of the US population on age, sex, race, education, region, metropolitan residence, veteran status and to account for oversamples and non-response.

likely than the youngest veterans to use the internet (OR = 0.03, CI = 0.01–0.06). Among internet using veterans there was no effect of age on health-related internet use. Higher income was positively associated with internet use (OR = 3.12, CI = 2.10–4.63, highest versus lowest income groups), but, like age, among internet using veterans income was not associated with health-related internet use. More education was associated with greater likelihood of internet use (OR = 4.20, CI = 2.92–6.02, most versus least educated group), and among internet using veterans it was associated with greater likelihood of health-related internet use (OR = 2.32, CI = 1.45–3.74, most versus least educated group). Health status had opposite effects in the two regression models. Worse health was associated with a smaller likelihood of general internet use (OR = 0.59, CI = 0.42–0.83, for fair/poor vs very good/excellent), but among internet using veterans, worse health was associated with greater likelihood of health-related internet use (OR = 1.85, CI = 1.16–2.95, for fair/poor vs very good/excellent health). Urban location was associated only with health-related internet use (OR = 2.41, CI = 1.66–3.50).

Discussion

Principal findings

This study represents the first large-scale, systematic report of veterans' use of the internet for health. It sheds light on socio-demographic characteristics that may substantially affect health organisations' ability to provide electronic, health-related information and services to patients with lower education levels and/or living in rural locations. We found that 29% of veterans had used the internet to search for health related information in the past year. Correlates differed between general internet use and health-related internet use. Younger age, higher income, more education and better health status were positively associated with general internet use, while education, living in urban areas and worse health status were positively associated with health-related internet use. The findings about education and rural location are each noteworthy. Education level was a barrier in two ways. First, those with less education were less likely to use the internet in general, and second, even among those who were internet users, less education was associated with less likelihood of using internet for health. Secondly, our study is among the first to document the importance of rural location on internet use for health, even after controlling for internet access and multiple socio-demographic variables. The potential rural barrier is particularly relevant today as the VA strives to improve access and quality of care for rural veterans.[22–24]

Implications of the findings

Our findings demonstrate that patient access to the internet does not guarantee their use of the internet for health-related information and services. Unless health-care organisations make special efforts to include disadvantaged groups, such as the less-educated and those in rural areas, the digital divide will persist. As organisations increasingly digitise they may inadvertently

erect barriers for some groups of vulnerable patients. This suggests that outreach may be needed. Healthcare organisations that intend to reach patients through the internet could provide direct training to vulnerable patients (e.g. those in rural areas, the less educated and those with stigmatised health conditions) to provide them with PHR and internet skills. Healthcare organisations should also consider sensitising and educating clinicians regarding patient internet use, and encourage clinicians to ask their patients if they use the internet for health. For patients who do not, clinicians could motivate them to do so by providing them with the URLs of informative, trustworthy and user-friendly health-related websites. Our findings about education suggest that website developers should carefully consider literacy and health literacy when designing content and layout. Less educated patients may be left behind as healthcare organisations digitise unless health-related websites are tailored to meet the needs of less literate patients. Finally, these data suggest that the VA's special focus on rural veterans is well placed.

Comparison with the literature

Our findings have similarities to, but also differences from, prior studies of internet use. General US population studies have shown that being female, being younger and having more education are positively associated with health-related internet use.[18,25] A recent study of users of a web-based personal health record found that white race and higher income were associated with greater use of the system.[26] Rice (2006) found that being female, being in worse health and having chronic health conditions were associated with health-related internet use.[9] Our results for veterans were similar but with two important distinctions. After controlling for being an internet user (i.e. analysis among veterans who were using the internet prior to the study start) age was not a statistically significant correlate of health-related internet use. Additionally, among the internet using veterans we found that rural respondents were less likely to use the internet for health. This is noteworthy because prior research shows that rural veterans have less contact with the VA, and worse health outcomes, than urban veterans.[22,27] Rural residents' use of the internet for health is an understudied area. Studies to date have been limited by their use of patient populations,[28] small scale qualitative designs[29] and use of convenience samples.[11,30]

Limitations of the method

Our cross-sectional study design precluded inferences of causality. Additionally, the data were collected in 2001 and 2002, so given the rapid development and adoption of online technologies more recent data are clearly desirable. However, because there has been little systematic research to date on veterans' use of the internet and use of the internet for health purposes, this study provides important insights about the socio-demographic characteristics that were barriers at the time of the study, and are likely to still be barriers today, though potentially attenuated. Even the most recent general population studies show that socio-demographic characteristics are still important determinants of health-related internet use.[10] Our results provide important additions to the knowledge about how socio-demographic characteristics affect electronic access to health information.

Call for further research

New surveys of veterans' use of the internet are needed to see if strides have been made in narrowing the gaps in health-related internet use based on education and rural–urban location. Additionally, to encourage greater use, research is needed to better understand physicians' views of patient use of the internet and PHRs. Such research could help healthcare organisations to enlist physicians to encourage their patients to use the internet and PHRs.

Our results provide a baseline against which future research can be compared, to determine whether gains have been made in providing health information and health services through the internet to a greater percentage of veterans. Future research will also evaluate whether disparities have been reduced, especially those based on education and urban–rural location. Additionally, research should examine the possibility that educational and urban–rural differences in internet use are not related to disparities in health information and health outcomes, but that they merely signify that different socio-demographic groups use different (but equally good) health information sources. Finally, additional research is needed to understand what aspects of rural locations might explain the lower rates of health-related internet use. It may be that innovations, such as the positive attributes of the internet for health information, do not diffuse as rapidly in rural locations as in urban locations due to less face-to-face contact.

Conclusions

Younger age, higher income and more education were associated with greater internet use. Among those who used the internet, using it for health-related purposes was strongly associated with education level and with

urban location, even after controlling for multiple socio-demographic characteristics. Healthcare facilities and systems need to recognise that there is unequal use of the internet for health, and that some groups will continue to be disadvantaged in terms of information and online services unless measures are taken to include such groups by tailoring services to their needs, providing training and tutorials and encouraging clinicians to learn about their patients' use of internet.

REFERENCES

1 Ayers SL and Kronenfeld JJ. Chronic illness and health-seeking information on the internet. *Health (London)* 2007;11:327–47.
2 Falvo DR. *Effective Patient Education*. Rockville, MD: Aspen Systems, 1985.
3 Rankin SH and Stallings KD. *Patient Education: issues, principles and practices* (2e). Philadelphia, PA: JB Lippincott, 1990.
4 McCaw B, McGlade K and McElnay J. The impact of the internet on the practice of general practitioners and community pharmacists in Northern Ireland. *Informatics in Primary Care* 2007;15:231–7.
5 Silvestre AL, Sue VM and Allen JY. If you build it, will they come? The Kaiser Permanente model of online health care. *Health Affairs (Millwood)* 2009;28:334–44.
6 Fox S. *Degrees of Access*. www.pewinternet.org/Commentary/2008/July/ (accessed 20 July 2009).
7 Harris Interactive. *Four Out of Five Adults Now Use the Internet*. 2008. www.harrisinteractive.com/harris_poll/ (accessed 17 July 2009).
8 Weaver JB 3rd, Mays D, Lindner G, Eroglu D, Fridinger F and Bernhardt JM. Profiling characteristics of internet medical information users. *Journal of the American Medical Informatics Association* 2009;16:714–22.
9 Rice RE. Influences, usage, and outcomes of internet health information searching: multivariate results from the Pew surveys. *International Journal of Medical Informatics* 2006;75:8–28.
10 Roblin DW, Houston TK 2nd, Allison JJ, Joski PJ and Becker ER. Disparities in use of a personal health record in a managed care organization. *Journal of the American Medical Informatics Association* 2009;16:683–9.
11 Zhang Y, Jones B, Spalding M, Young R and Ragain M. Use of the internet for health information among primary care patients in rural West Texas. *Southern Medical Journal* 2009;102:595–601.
12 Halamka JD, Mandl KD and Tang PC. Early experiences with personal health records. *Journal of the American Medical Informatics Association* 2008;15:1–7.
13 Nazi KM and Woods SS. MyHealtheVet PHR: a description of users and patient portal use. *American Medical Informatics Association Annual Symposium Proceedings* 2008:1182.
14 Gordon KS, Brandt CA, Goulet JL and Justice AC. *Veterans' Access to Internet Varies by Race and Age: potential impact on VHA electronic health communication efforts*. Health Services Research and Development National Meeting, Baltimore, MD, 11 to 13 February 2008.
15 Schneiderman AI, Lincoln AE, Curbow B and Kang HK. Variations in health communication needs among combat veterans. *American Journal of Public Health* 2004;94:2074–6.
16 Lorence DP, Park H and Fox S. Racial disparities in health information access: resilience of the digital divide. *Journal of Medical Systems* 2006;30:241–9.
17 Department of Veterans Affairs Public and Intergovernmental Affairs. *VA Announces $22 Million for Rural Veterans*. Washington, DC: VA, 2009. www1.va.gov/opa/pressrel/pressrelease.cfm?id=1642> (accessed 23 November 2009).
18 Baker L, Wagner TH, Singer S and Bundorf MK. Use of the internet and e-mail for health care information: results from a national survey. *Journal of the American Medical Association* 2003;289:2400–6.
19 Bundorf MK, Wagner TH, Singer SJ and Baker LC. Who searches the internet for health information? *Health Services Research* 2006;41:819–36.
20 Berger M, Wagner TH and Baker LC. Internet use and stigmatized illness. *Social Science and Medicine* 2005;61:1821–7.
21 Wagner TH, Baker LC, Bundorf MK and Singer S. Use of the internet for health information by the chronically ill. *Preventing Chronic Disease* 2004;1:A13.
22 Wallace AE, Weeks WB, Wang S, Lee AF and Kazis LE. Rural and urban disparities in health related quality of life among veterans with psychiatric disorders. *Psychiatric Services* 2006;57:851–6.
23 Mooney C, Zwanziger J, Phibbs CS and Schmitt S. Is travel distance a barrier to veterans' use of VA hospitals for medical surgical care? *Social Science and Medicine* 2000;50:1743–55.
24 Weeks WB, Kazis LE, Shen Y *et al*. Differences in health related quality of life in rural and urban veterans. *American Journal of Public Health* 2004;94:1762–7.
25 Satterlund MJ, McCaul KD and Sandgren AK. Information gathering over time by breast cancer patients. *Journal of Medical Internet Research* 2003;5:e15.
26 Schnipper JL, Gandhi TK, Wald JS *et al*. Design and implementation of a web-based patient portal linked to an electronic health record designed to improve medication safety: the Patient Gateway medications module. *Informatics in Primary Care* 2008;16:147–55.
27 Weeks WB, Wallace AE, Wang S, Lee A and Kazis LE. Rural–urban disparities in health related quality of life within disease categories of Veterans. *Journal of Rural Health* 2006;22:204–11.
28 Abdullah M, Theobald DE, Butler D *et al*. Access to communication technologies in a sample of cancer patients: an urban and rural survey. *BMC Cancer* 2005;5:18.
29 Tillotson S, Lear S, Araki Y *et al*. Innovation in the North: are health service providers ready for the uptake of an internet-based chronic disease management platform? *Studies in Health Technology and Informatics* 2009;143:472–7.
30 Quin J, Stams V, Phelps B, Boley T and Hazelrigg S. Interest in internet lung cancer support among rural cardiothoracic patients. *Journal of Surgical Research* 2008;144:275–6.

CONFLICTS OF INTEREST

None.

ADDRESS FOR CORRESPONDENCE

D Keith McInnes
Center for Health Quality, Outcomes and Economic Research
VA Bedford Medical Center
200 Springs Road (152)
Bedford MA 01730
USA
Email: keith.mcinnes@va.gov

Accepted February 2010

Program Characteristics Associated With Testing for HIV and Hepatitis C in Veterans Substance Use Disorder Clinics

D. Keith McInnes, Sc.D., M.S.
Jenny K. Hyun, Ph.D., M.P.H.
Jodie A. Trafton, Ph.D.
Steven M. Asch, M.D., M.P.H.
Allen L. Gifford, M.D.

Objective: **This study examined whether organizational characteristics and quality improvement initiatives were related to HIV and hepatitis C (HCV) testing rates in veterans' substance use disorders programs.** ***Methods:*** **Data were collected by surveying 232 program directors at all U.S. Department of Veterans Affairs (VA) substance use disorder programs.** ***Results:*** **Program directors (N=223) reported that, on average, 35% of their patients were tested for HIV (median=10%) and 57% were tested for HCV (median=80%). Of the quality improvement initiatives examined, computerized reminders to clinicians (p=.02) and a designated clinician for screening (p=.01) were positively associated with HCV testing, and computerized templates that guide clinicians through ordering of testing were positively associated with HIV testing (p=.06).** ***Conclusions:*** **Despite national emphasis on HIV testing, rates of testing were lower for HIV than for HCV in programs serving veterans with substance use disorders and at risk of both illnesses. System-level quality improvement initiatives may be effective at increasing rates of infectious disease screening. (*Psychiatric Services* 61:90–94, 2010)**

Dr. McInnes is affiliated with the Center for Health Quality Outcomes and Economic Research (CHQOER), Edith Nourse Rogers Memorial Veterans Hospital, 200 Springs Rd. (152), Bldg. 70, Bedford, MA 01730 (e-mail: keith.mcinnes@va.gov). He is also with the Department of Health Policy and Management, Boston University School of Public Health, Boston. Dr. Hyun is with the National Center for PTSD, Department of Veterans Affairs (VA) Palo Alto, Palo Alto, California. Dr. Trafton is with the VA Center for Health Care Evaluation, Menlo Park, California. Dr. Asch is with the HIV Hepatitis Quality Enhancement Research Initiative, VA Greater Los Angeles Health Care System, and the Department of Medicine, David Geffen School of Medicine at the University of California, Los Angeles. Dr. Gifford is with the HIV Hepatitis Quality Enhancement Research Initiative, VA New England Health Care System, Bedford, Massachusetts, and Boston University Schools of Public Health and Medicine.

HIV and hepatitis C virus (HCV) infections are leading causes of death and disability in the United States. Much of the morbidity and mortality from these conditions is due to late detection and late entry into care (1,2), which could be addressed by more widespread screening and testing. The Department of Veterans Affairs (VA) is the largest provider of HIV and HCV care in the United States. Despite efforts to increase testing among veterans (3,4), rates of HIV and HCV testing remain suboptimal, even among those with known risk factors. Substance use disorders increase the risk of infection by HCV (5) and HIV (6,7). Despite this increased risk, a recent study in the Pacific Northwest found that only 19% of VA patients with substance use disorders had been tested for HIV and only 60% had been tested for HCV (8).

Beginning in 1999 the VA mandated its health facilities to conduct universal screening for HCV risk factors (and to conduct antibody testing if risk factors were present), created a computerized clinical reminder in the electronic medical record, which prompts providers to screen and test their patients for HCV, and made HCV risk screening a quality performance measure (3). The VA has also devoted resources to increasing HIV screening rates (4). Recent research has shown that quality improvement initiatives such as provider feedback and provider activation can increase HIV testing rates among at-risk patients (4). Little is known, however, about how these policies and characteristics of substance use disorder programs may affect infectious disease testing rates in substance use disorder programs.

The Drug and Alcohol Program Survey (DAPS) collects data about the VA's substance use disorder programs, which treat over 120,000 patients annually. In 2006 all VA substance use disorder clinics (N=232) responded to the DAPS survey, which for the first time included questions about HIV and HCV testing and about system-level initiatives to increase testing rates (for example, computerized reminders and provider feedback). We describe

HIV and HCV testing rates by substance use disorder program type. We also evaluate whether quality improvement systems are related to HIV and HCV testing rates.

Methods

We surveyed every VA program that was designed to treat patients who have substance use disorders, had at least two full-time staff, and could be distinguished from other programs on the basis of unique staffing, patients, clinical services, or policies. The survey response rate was 100% (N=232). Of these, nine programs did not meet the above inclusion criteria, resulting in 223 programs in our analyses.

The survey was mandated in all VA substance use disorder programs, was conducted between October 1, 2006, and January 31, 2007, and was completed by the substance use disorder program director or designee. It covered patient characteristics, treatments, staffing, continuity of care, HIV and HCV screening and testing, and quality improvement systems in use. For most items, respondents were asked to describe programs and characteristics in place during September 2006. Program directors were asked about services directly delivered within their substance use disorder programs. Information gathered included the outcome measures of the percentage of clinic patients who, between October 1, 2005, and September 30, 2006, received HIV testing and HCV testing from program staff. We combined conceptually related items into indices.

The six quality improvement systems assessed were provider education, computerized reminders, computerized templates to guide test orders, provider performance profiling and feedback, presence of a clinical champion to promote and monitor the progress of testing, and presence of a designated clinician for testing. All of these variables were dichotomous. They were selected because they are commonly used quality improvement methods that have been used to increase infectious disease testing (4,9,10).

Treatments and services provided in the substance use disorder programs were assessed through two variables. Supplementary services referred to smoking cessation and vocational training, and the variable was calculated by summing the percentages of patients engaged in a smoking cessation program or in vocational training and then dichotomizing at the median (a code of 1 indicated more patients receiving services). Evidence-based medication therapy was a dichotomous variable, where 1 indicated use of any of five treatments (methadone, buprenorphine, or naltrexone for opioid use, and naltrexone or acamprosate for alcohol use) and 0 indicated that none of these treatments were used.

Six variables described program characteristics, such as program type (such as inpatient or residential), staffing, patients, and emphasis on infectious disease testing. The priority a clinic places on infectious disease testing was assessed by asking respondents to compare HIV testing (for the first regression model) and HCV testing (for the second regression model) with the "assessment and treatment of other common conditions (mental health care, diabetes monitoring, cancer screening, chronic pain management)," scored as 0, not important; 1, somewhat important; and 2, very important. Substance use disorder programs were uniquely assigned to one category according to type of treatment and services they provided (11). Inpatient programs provided 24-hour acute care. Residential programs provided round-the-clock care but without 24-hour medical staff. Intensive outpatient programs provided at least three hours of programming three days per week, and standard outpatient programs (the reference category in multivariate models) provided care that did not meet the frequency criteria of intensive treatment for outpatients. All programs provided psychosocial rehabilitation.

Medical staff was assessed by the number of physicians, physician assistants, and nurses divided by total staff and then dichotomized at the median; social welfare staff was quantified as the sum of psychologists, social workers, addiction therapists, and psychiatric technicians or aides divided by total staff and then dichotomized at the median (with 1 indicating more staff). Number of unique patients represented patients admitted to the program during the 2006 fiscal year. Patients with complex needs represented five summed variables—percentages of patients who were homeless (or in unstable living arrangements), nicotine dependent, opioid dependent, or not married (or in a long-term relationship) or who had a major psychiatric disorder, with results dichotomized at the median (with 1 indicating more patients with complex needs).

Descriptive statistics were estimated for total respondents and for each program type (inpatient, residential, intensive outpatient, or standard outpatient). We compared program types by organizational characteristics by using analysis of variance and chi square tests. Next, we estimated differences in the use of quality improvement systems for HIV testing versus HCV testing using chi square tests.

Multivariate regression analyses were conducted with HIV and HCV testing rates as the dependent variables. In the models, we retained variables that in bivariate analyses were significantly associated ($p \leq .05$) with either HIV testing or HCV testing. In addition, despite bivariate nonsignificance, for conceptual reasons we retained clinical champion, inpatient program, intensive outpatient program, and number of patients in multivariate models. We examined interactions between designated clinician for screening and medical staffing levels, between designated clinician for screening and number of patients, and between evidence-based medication treatments and medical staffing levels. We conducted sensitivity analyses to see whether our results differed when the outcome variables (HIV and HCV testing rates) were dichotomous instead of continuous. We used three cut points for dichotomization: 75%, 50%, and 25% of patients tested. This study was approved by the Stanford University Institutional Review Board and Veterans Affairs Palo Alto Health Care System Research and Development Committee.

Results

In all programs combined (N=223), mean±SD reported testing rates were lower for HIV (35%±40% of patients in the program, median 10%) than for HCV (57%±44%, median 80%) (t=8.61, df=222, p<.001 for difference of means). Residential programs (N=65) reported testing the largest proportion of patients for HIV and HCV. For HIV, residential programs tested 51%±44% of patients, versus 23%±35% for outpatient programs (N=97) (F=6.52, df=3 and 219, p<.001). For HCV, residential programs tested 73%±39% of patients, versus 45%±45% for outpatient programs (F=5.63, df=3 and 219, p=.001). Inpatient programs (N=19) tested 40%±42% and 63%±43% of patients, on average, for HIV and HCV, respectively, whereas intensive outpatient programs (N=42) tested 36%±39% and 55%±44% of patients, respectively. About one-third of programs (N=77 or 35%) tested more than half of their patients for HIV, 131 programs (59%) tested more than half for HCV, and 74 programs (33%) tested more than half of their patients for both HIV and HCV. The Pearson correlation between HIV and HCV testing was .61 (p<.01). Residential programs, compared with the other types of program, rated the importance of HIV testing (F=5.29, df=3 and 217, p=.002) and HCV testing (F=4.75, df=3 and 216, p=.003) highest compared with testing for other conditions, and outpatient and inpatient programs rated the importance of HIV and HCV testing lowest.

On average, of the quality improvement systems, programs used computerized reminders most frequently (91 programs, 41%, for HIV and 164 programs, 74%, for HCV) (χ^2=22.28, df=1, p<.01), followed by provider education (122 programs, 55%, for HIV versus 131 programs, 59%, for HCV), computerized templates (92 programs, 41%, for HIV versus 106 programs, 48%, for HCV), designated clinician for testing (78 programs, 35%, for HIV versus 84 programs, 38%, for HCV), performance profiling (32 programs, 14%, for HIV versus 41 programs, 18%, for HCV), and clinical champion for testing (19 programs, 9%, for HIV versus 34 programs, 15%, for HCV). These systems were used more for HCV than for HIV, although the difference was significant for computerized reminders only.

Results of multivariate models are shown in Table 1. Computerized templates were positively associated with HIV testing, although the association was not quite significant (p=.06). For HCV testing, use of computerized reminders (p=.02) and use of a designated clinician for screening (p=.01) were significant. Provision of supplementary services (smoking cessation services and vocational training) was positively associated with HCV testing (p=.001), and use of evidence-based medication therapy was positively associated with both HIV testing (p=.01) and HCV testing (p<.001). Among program characteristics, priority placed on infectious disease testing was significantly associated with HCV testing (p=.002), residential programs (compared with standard outpatient programs) were more likely to test for both HIV (p<.001) and HCV (p=.01), and the variable patients with complex needs was positively associated with HIV testing (p=.03). None of the interactions we examined (between designated clini-

Table 1

Characteristics associated with HIV and hepatitis C testing in 223 VA substance use disorder programs[a]

Characteristic	HIV testing[b]			Hepatitis C testing[c]		
	Estimate	SE	p	Estimate	SE	p
System-level initiative to increase testing						
Provider education	–1.93	5.52	.73	2.21	5.45	.69
Computerized reminders for screening	10.07	5.84	.09	14.51	6.31	.02
Computerized templates for screening	10.65	5.65	.06	5.35	5.51	.33
Provider profiling and feedback for screening	2.07	7.88	.79	8.52	6.98	.22
Clinical champion for screening	–9.75	9.44	.30	–7.96	7.69	.30
Designated clinician for screening	9.78	5.84	.10	14.48	5.87	.01
Treatments and services						
Supplementary services	3.28	5.41	.54	18.47	5.33	.001
Evidence-based medication therapy	16.03	5.68	.01	24.85	5.64	<.001
Program characteristics						
Priority placed on infectious disease testing[d]	8.02	5.12	.12	18.68	5.91	.002
Inpatient program[e]	10.53	10.38	.31	11.66	10.37	.26
Residential program[e]	22.8	6.79	<.001	17.76	6.61	.01
Intensive outpatient program[e]	12.42	7.49	.10	7.68	7.22	.29
Medical staff	–3.89	6.39	.54	–2.78	6.22	.65
Social welfare staff	–2.62	6.31	.68	–7.94	6.14	.20
Number of unique patients	.00	.00	.74	.00	.00	.50
Patients with complex needs	11.67	5.38	.03	4.76	5.41	.38

[a] Analysis of testing rates in the Veterans Affairs (VA) system was by multivariate ordinary least-squares regression.
[b] Adjusted R^2=.16, F=3.51, df=16 and 200, p<.001
[c] Adjusted R^2=.32, F=7.42, df=16 and 199, p<.001
[d] Separate items were used for program priority for HIV testing (used in the HIV testing model) and for HCV testing (used in the HCV testing model).
[e] Reference group is standard outpatient setting.

cian and medical staffing levels, between designated clinician and number of patients, and between evidence-based medication treatments and medical staffing levels) were statistically significant. Our sensitivity analyses with dichotomous outcome variables produced substantially the same results as with continuous outcomes.

Discussion

In the first nationwide survey of HIV and HCV testing in VA substance use disorder programs, 223 program directors reported testing, on average, 35% of patients for HIV and 57% for HCV. The highest rates of testing were in residential programs (51% for HIV and 73% for HCV), and the lowest rates were in standard outpatient programs (23% for HIV and 45% for HCV). Computerized templates to guide test orders were associated with higher reported HIV testing rates, although the relationship did not achieve statistical significance (p=.06). Computerized reminders and designated clinician for screening were associated with higher reported HCV testing rates. There was substantial variation in the use of the quality improvement systems, ranging from computerized reminders used by 91 (41%) and 164 (74%) programs, respectively for HIV and HCV testing, to clinical champions for testing used by 19 (9%) and 34 (15%) programs, respectively.

Reported testing rates for both HIV and HCV are strikingly low, especially given that patients with substance use disorders are at high risk for both conditions and that guidelines supporting testing of these groups are widely endorsed and disseminated. There are several possible explanations for low testing rates. Providers with many demands may focus on their immediate job of treating substance use disorders. Especially with opioid-dependent patients, providers may feel a sense of urgency to stabilize their patients above all else, in order to prevent relapse and possible death from overdose. Other providers may not be aware of the importance of viral screening. Also, patients' refusal of testing could contribute to low testing rates.

The 22% gap we found between mean reported HIV and HCV testing rates may reflect HIV testing policies in the VA at the time that made HIV testing more of a clinical burden than other blood tests by requiring written informed consent and pre- and posttest counseling. Also, until recently, VA policy strongly emphasized HCV testing by mandating computerized clinical HCV testing reminders. Also, the greater prevalence of HCV infection among veterans may lead to greater provider awareness and higher priority for HCV testing and treatment. The positive association between percentage of patients with complex needs and HIV testing rates may similarly reflect greater interest in infectious disease screening when clinicians and managers perceive, on the basis of patient characteristics, that HIV prevalence may be high in their patient population.

Of interest is that use of evidence-based medication therapies for opioid and alcohol use was significantly associated with both HIV and HCV testing rates. We suspect that programs using evidence-based medication therapies are more medically (versus behaviorally) oriented and thus may order more medical diagnostic tests, such as those for HIV and HCV. Also noteworthy, computerized templates were positively associated with only HIV testing, whereas computerized reminders were significantly associated with only HCV testing. The former may indicate that computerized templates were well suited for HIV testing in the VA because of the multiple steps and forms required when we conducted our survey (the process in the VA has been streamlined since then), and the latter may reflect the VA's implementation of the HCV testing reminder in the electronic medical record (3). The association of a designated clinician for HCV screening may indicate that providers are more likely to order tests if they know that a responsible clinician will ensure their implementation.

Our study has several limitations. Data were collected cross-sectionally, which limits our ability to make causal inferences. We used reports from program directors to assess HCV and HIV testing rates. The rates we observed may be subject to reporting bias (such as inflated estimates). Our findings for HIV testing, however, are similar to recent, separate analyses of VA administrative data that suggest that about 28% of veterans treated in substance use disorder programs receive HIV testing (12). Thus it is likely that the rates of testing shown here, although generally low, may have been even lower if we used different methods. Our independent variables were also based on respondent report and were not independently verified. It is important to note, however, that substance use disorder programs are the only repository of information about services and testing provided to patients in those programs. VA administrative data structures do not allow unambiguous matching of patients to specific substance use disorder programs. Thus this survey represented the most efficient means of examining system-level issues in substance use disorder programs and enabled the comparison of program-level testing rates with program-level characteristics, such as services provided and staffing levels.

Conclusions

Despite limitations, these national survey results suggest strikingly low rates of testing in VA substance use disorder programs, with 35% of patients tested for HIV and 57% tested for HCV. The gap between the rates may reflect differences in VA policies (at the time of the survey, written informed consent was required for HIV testing) or greater awareness, among providers, of HCV because of its higher prevalence. Our finding of an association between quality improvement systems and higher testing rates is encouraging for substance use disorder programs interested in increasing testing rates. Most promising were computerized templates for HIV testing and computerized reminders and designated clinicians for HCV testing.

Acknowledgments and disclosures

This research was supported by the VA Quality Enhancement Research Initiative HIV/hepatitis C; VA Bedford Center for Health Quality, Outcomes and Economic Research; VA Palo Alto Center for Health Care Evaluation; and the VA Palo Alto National Center for PTSD.

The authors report no competing interests.

References

1. Mallette C, Flynn MA, Promrat K: Outcome of screening for hepatitis C virus infection based on risk factors. American Journal of Gastroenterology 103:131–137, 2008
2. Caring for Veterans With HIV Disease: Characteristics of Veterans in VA Care. Washington, DC, US Department of Veterans Affairs, Public Health Strategic Health Care Group, Center for Quality Management in Public Health, 2003
3. Huckans MS, Blackwell AD, Harms TA, et al: Management of hepatitis C disease among VA patients with schizophrenia and substance use disorders. Psychiatric Services 57:403–406, 2006
4. Goetz MB, Bowman C, Hoang T, et al: Implementing and evaluating a regional strategy to improve testing rates in VA patients at risk for HIV utilizing the QUERI process as a guiding framework. Implementation Science 3:16, 2008
5. National Institutes of Health Consensus Development Conference Statement: management of hepatitis C. Hepatology 36:S3–S20, 2002
6. Rees V, Saitz R, Horton NJ, et al: Association of alcohol consumption with HIV sex- and drug-risk behaviors among drug users. Journal of Substance Abuse Treatment 21:129–134, 2001
7. Strathdee SA, Sherman SG: The role of sexual transmission of HIV infection among injection and non-injection drug users. Journal of Urban Health 80:iii,7–14, 2003
8. Huckans MS, Blackwell AD, Harms TA, et al: Integrated hepatitis C virus treatment: addressing comorbid substance use disorders and HIV infection. AIDS 19:S106–S115, 2005
9. Renders CM, Valk GD, Griffin S, et al: Interventions to improve the management of diabetes mellitus in primary care, outpatient and community settings. Cochrane Database of Systematic Reviews 4:CD001481, 2001
10. Wagner EH, Austin BT, Davis C, et al: Improving chronic illness care: translating evidence into action. Health Affairs 20(6): 64–78, 2001
11. Tracy SW, Trafton JA, Humphreys K: The Department of Veterans Affairs Substance Abuse Treatment System: Results of the 2003 DAPS. Washington, DC, US Department of Veterans Affairs, 2004
12. Dookeran NM, Burgess JF, Bowman C, et al: HIV screening among substance abusing veterans in care. Journal of Substance Abuse Treatment 37:286–291, 2009

Are Residents' Decisions Influenced More by a Decision Aid or a Specialist's Opinion? A Randomized Controlled Trial

Mitchell A. Medow, MD, PhD[1], Hal R. Arkes, PhD[2], and Victoria A. Shaffer, PhD[3]

[1]Section of General Internal Medicine, Boston University School of Medicine, Boston, MA, USA; [2]Department of Psychology, The Ohio State University, Columbus, OH, USA; [3]Department of Psychology, Wichita State University, Wichita, KS, USA.

BACKGROUND: Physicians are reluctant to use decision aids despite their ability to improve care. A potential reason may be that physicians do not believe decision aid advice.

OBJECTIVE: To determine whether internal medicine residents lend more credence to contradictory decision aid or human advice.

DESIGN: Randomized controlled trial. Residents read a scenario of a patient with community-acquired pneumonia and were asked whether they would admit the patient to the intensive care unit or the floor. Residents were randomized to receive contrary advice from either a referenced decision aid or an anonymous pulmonologist. They were then asked, in light of this new information, where they would admit the patient.

PARTICIPANTS: One hundred eight internal medicine residents.

MEASUREMENTS: The percentage of residents who changed their admission location and the change in confidence in the decision.

MAIN RESULTS: Residents were more likely to change their original admission location (OR 2.3, 95% CI 1.04 to 5.1, P=0.04) and to reduce their confidence in the decision (adjusted difference between means −12.9%, 95% CI −3.0% to −22.8%, P=0.011) in response to the referenced decision aid than to the anonymous pulmonologist. Confidence in their decision was more likely to change if they initially chose to admit the patient to the floor.

CONCLUSIONS: In a hypothetical case of community-acquired pneumonia, physicians were influenced more by contrary advice from a referenced decision aid than an anonymous specialist. Whether this holds for advice from a respected specialist or in actual practice remains to be studied.

KEY WORDS: decision aid; judgment; choice behavior; internship and residency.

J Gen Intern Med 25(4):316–20
DOI: 10.1007/s11606-010-1251-y

Physicians are reluctant to use medical decision aids, despite their documented ability to improve the quality of patient care[1–6]. For example, Garg's systematic review[1] found that clinical decision support systems improved performance in 64% of the 97 studies identified. Yet physicians do not use decision aids. Only 4% of physicians in a recent national survey of ambulatory physicians had access to a fully functional electronic health record that offered rudimentary decision support (medication warnings, out-of-range laboratory indications and guideline-based reminders).[7] Even physicians exposed to a helpful aid in a randomized controlled trial and who experienced its benefits did not continue to use it after the trial [8]. Factors relating to this underutilization are commonly classified as pertaining to the clinical problem, patient factors, provider traits, system characteristics and the health-care environment [9]. We have been exploring issues affecting the psychological perspective of the provider on the use of decision aids [10,11]. Perhaps physicians simply do not put as much credence in the recommendation of a decision aid compared with the advice of a colleague. Dreiseit and Binder [12] found that dermatologists changed their decisions in only a quarter of the cases when a decision aid contradicted them. To ensure conflict between the physicians and the decision aid, their system purposefully gave an incorrect diagnosis in about a third of the cases presented.

To further test the hypothesis that physicians may not believe decision aids as much as they believe colleagues, in the present study we gave internal medicine resident physicians a case scenario of a patient with pneumonia and asked them whether they would admit the patient to the floor or the intensive care unit (ICU). We then presented them with contradictory advice from either a referenced decision aid or an anonymous pulmonologist and again asked, in light of the new advice, to which location they would admit the patient. No subterfuge was needed in this scenario—two validated decision aids made different recommendations based on the same data.

Presented at the 31st Annual Meeting of the Society for Medical Decision Making, October 18–21, 2009, and the 30th Annual Conference of the Society for Judgment and Decision Making, November 21–23, 2009.

Support: The Program in Decision, Risk, and Management Science at the National Science Foundation (NSF).

Electronic supplementary material *The online version of this article (doi:10.1007/s11606-010-1251-y) contains supplementary material, which is available to authorized users.*

Received December 2, 2008
Revised May 27, 2009
Accepted December 21, 2009
Published online January 30, 2010

METHODS

Research Question

Do physicians, when confronted with contradictory advice, lend more credence to a referenced decision aid or an anonymous specialist?

Subjects

One hundred eight residents in The Ohio State University (OSU) Medical Center's internal medicine residency program (preliminary, categorical, combined Medicine/Pediatrics and chief residents) were recruited in November 2007. Subjects were recruited through a written letter and an oral request at a conference. They were e-mailed occasional reminders if they did not complete the survey and a thank-you note when they completed the survey. To augment our enrollment we used a convenience sample of internal medicine residents from MetroHealth Medical Center, recruited by one of their chief residents with a single solicitation in June 2008. They completed the survey in July. All participants were paid $50.

Procedure

The study was conducted electronically over the World Wide Web. Each resident was presented the case of community-acquired pneumonia (CAP) and asked if the patient should be admitted to the floor or the ICU (figure in online-only Appendix). The case was carefully constructed such that the two major decision aids for CAP offer different recommendations. The PORT (Pneumonia Patient Outcomes Research Team) Score[13] suggests a floor admission, while the CURB-65 (Confusion, Urea, Respiratory Rate, Blood Pressure and age 65 or older)[14] suggests an ICU admission. After reading the case subjects were asked:

(a) Do you feel that Ms. J should be admitted to the intensive care unit (ICU) or to the floor? (Please choose one option.)
(b) Please indicate your confidence in your decision by selecting a level below from 50% (no confidence) to 100% (complete confidence).

After answering these two questions, subjects were then presented with contradictory advice. For example, if they decided to admit the patient to the floor, they were presented with advice to admit the patient to the ICU. Each subject was randomized by computer to get this contradictory advice from either (a) a referenced decision aid (DA) or (b) an anonymous pulmonologist.

An example of the advice from the decision aid is "A severity score has been developed for community acquired pneumonia (Fine MJ, Auble TE, Yealy DM, et al., "A Prediction Rule to Identify Low-Risk Patients with Community-Acquired Pneumonia." N Engl J Med. 1997 Jan 23;336 (4):243–50). The community-acquired severity score classifies this case as Class IV, for which hospital admission is advised, but ICU admission is not necessary."

An example of the advice from the anonymous pulmonologist is, "As you are getting ready to admit the patient, you run into a pulmonologist and decide to ask their advice. After you present the patient, you ask the pulmonologist whether Ms. J should be admitted to the ICU or to the floor. The pulmonologist suggests a floor admission."

After reading the contradictory advice, the subjects were again asked whether the patient should be admitted to the intensive care unit (ICU) or to the floor, and for their confidence in that decision.

Analysis

The primary endpoint was the comparison of the number of subjects that switched their decision about the admission location (e.g., floor to ICU) when given the decision aid or pulmonologist's advice (randomized), stratified by their initial decision (self-selected, i.e., floor or ICU). We used a logistic regression model to check the independent statistical significance of the change in the admission location with respect to (1) the source of the advice (randomized) and (2) the initial admission location (self-selected). We used a full model with all interaction terms among the initial admission location, source of advice and residency program to determine if any of the interaction terms were statistically significant.

The secondary endpoint was the change in confidence in the decision. Confidence change was calculated as the change in confidence between the final and initial confidence indicated by the subjects, adjusted for changes of location. The change in confidence was analyzed with an analysis of variance (ANOVA) model including the source of advice (randomized) and the initial admission location (self-selected). A full model, with all the interaction terms, was also tested.

All statistical tests were two-sided with α=0.05. The analysis was done with R version 2.9.2[15].

Sample Size

We would need 65 subjects in each group (130 total subjects) to have 90% power to detect a difference in the proportion of residents who changed their admission location from 5% to 25%, with a two-sided α of 0.05. We had enough incentive funds to survey 110 subjects; therefore, that was our target enrollment with 84% power to find the hypothesized difference.

Institutional Review and Funding

The study was approved by The Ohio State Institutional Review Board.

RESULTS

Subjects

Of the 108 residents solicited from OSU, 85 responded, 2 with incomplete data (77% response rate). Twenty-five of 99 solicited residents from MetroHealth completed the survey. Thus, there was a total of 108 participants. The subjects were mostly categorical internal medicine (71) and combined internal medicine-pediatrics (24) residents. Fifty-two subjects were randomized to receive contrary advice from the decision aid and 56 from the pulmonologist (Table 1).

Location Change

Overall 48 of the 108 subjects (44.4%, 95% CI: 34.9–54.3%) changed the admission location in response to the contrary

Table 1. Subject Demographics

	Source of contradictory advice	
	Specialist (n=56)	Decision aid (n=52)
Year of training		
1	14	18
2	21	8
3	16	18
4	5	7
5	0	1
Training program		
Internal medicine (preliminary)	5	4
Internal medicine (categorical)	34	37
Internal medicine–pediatrics	14	10
Other	3	1
Site		
MetroHealth	16	9
The Ohio State University	40	43

advice (Table 2). When the resident received advice from the decision aid, the resident changed the admission location in 53.8% of the cases (95% CI: 39.5–67.8%). When the resident received advice from the pulmonologist, the resident changed the admission location in only 35.7% of the cases (95% CI: 23.3–49.6%). The unadjusted relative risk for a subject to change their admission location was 1.51, indicating that subjects were approximately 50% more likely to change their decisions in response to advice from the decision aid than from a specialist. The unadjusted odds ratio (OR) was 2.10. The adjusted OR from the logistic regression model including the initial admission location (the pre-specified primary endpoint) was 2.27 (95% CI: 1.04–5.08; P=0.04), indicating that the odds of changing the admission location if given contradictory advice from the decision aid were over two times the odds of changing the decision if given advice from the pulmonologist. In the full logistic regression model with all the interaction terms included, none of the terms (the source of advice, the initial admission and the interaction terms) showed significance. Significance of the interaction terms would have suggested that the effect of one of the terms (e.g., the source of advice) had different effects depending on the value of the other term (e.g., the initial admission location). There was no association between location change and either initial confidence (P=0.50) or year of training (P=0.33).

Confidence Change

In the ANOVA model exploring changes in confidence, both the subjects' initial admission location and the source of advice were statistically significant. Respondents were more likely to change their confidence in their initial decision towards the contrary advice after hearing the decision aid's recommendation (unadjusted mean, −36.0%; SD=29.7) than the pulmonologist's recommendation (unadjusted mean, −23.0%; SD=22.9; adjusted difference between means, −12.9%; 95% CI –3.0% to −22.8%, P=0.011), irrespective of their initial admission decision. Second, residents were more likely to lower their confidence in their initial decision if their initial admission location was to the floor (unadjusted mean, −32.8%; SD=26.0) than to the ICU (unadjusted mean, −22.2%; SD=28.1; adjusted difference between means −12.0%, 95% CI −1.5% to −22.4%, P=0.025). In the full model including all the interaction terms, the main effects of source of advice and initial admission location were significant ($P \leq 0.025$), whereas the program and the interaction terms were not statistically significant ($P > 0.13$).

DISCUSSION

Contradictory advice from a referenced decision aid was more effective in influencing resident physician decisions than contradictory advice from an anonymous specialist, as evidenced by the change in admission location and the change in confidence. Interpretation of this result is complicated by the asymmetry in the description of the alternatives. In designing the study we elected to provide the minimal necessary description of the decision aid, which we felt to be identifying information (i.e., its name and literature citation), and its recommendation (e.g., "The community-acquired severity score classifies this case as Class IV, for which hospital admission is advised, but ICU admission is not necessary."). We purposefully did not repeat data used by the aid (e.g., respiratory rate) or the aid's test characteristics (e.g., positive predictive value). We kept the human opinion anonymous since locally respected leaders can affect adherence[16]. Instead, we vested the human in the mantle of a specialist. Thus, the description of the decision aid, although minimal, was more elaborate than that of the human, which could have affected the results. It is unclear if using an anonymous decision aid or a locally respected human would have produced the same results.

This result is unexpected given the general reluctance of physicians to utilize decision aids[17]. One might have surmised that physicians do not find decision aid advice credible. Yet the physicians in our study were more likely to be influenced by the decision aid. In the clinic physicians might still eschew decision aids because of other problems, such as patient, provider, system and health-care environment factors. By design, we did not examine whether the resident would use a decision aid, the residents' opinions of decision aids or the residents' attitudes about physicians who use decision aids. In a previous study, Arkes et al.[10] showed that patients and

Table 2. Location Change Results

	Initial admission to the ICU		Initial admission to the floor	
	Switched	Did not switch	Switched	Did not switch
Decision aid	9 (9/20=45%)	11 (11/20=55%)	19 (19/32=59%)	13 (13/32=41%)
Specialist	4 (4/16=25%)	12 (12/16=75%)	16 (16/40=40%)	24 (24/40=60%)

ICU = Intensive care unit. Odds ratio for the source of advice (decision aid vs. pulmonologist) 2.3, 95% CI 1.0 to 5.1, P=0.04. Odds ratio for the initial admission location (ICU vs. floor) 0.5, 95% CI 0.2 to 1.2, P=0.15

medical students derogate physicians who use decision aids. This suggests that physician attitudes about using decision aids and their actions when given (unsolicited) decision aid recommendations diverge.

Only about half of the physicians in training were willing to change their decisions based on advice from a more qualified source. We cannot conclude that the physicians in our study who maintained their initial hypothesis were mistakenly preserving an incorrect one. First a decision to admit to either the ICU or the floor was justified by one of the two reputable decision aids on this topic. Second, assessing the accuracy of management decisions, such as the decision to admit to ICU, is more difficult than assessing the accuracy of diagnostic decisions, as the criterion for accuracy is less definitive. Thus, management decisions may not have an indisputably correct course of action that should resist the contrary guidance of a decision aid or colleague.

Although our research focuses on a management decision, parallels drawn from the literature on diagnostic decisions may inform this discussion. Berner and colleagues[18] found that residents often maintained an unquestionably incorrect diagnosis despite a decision support system's suggestion of the correct diagnosis. They suggested that their findings may have been due to overconfidence on the part of the residents. In support of Berner and colleagues' suggestion, Dreiseit and Binder[12] found in a study of dermatologists' decision making that there was a negative correlation between a physician's confidence and their propensity to change their decisions. These studies suggest that if a physician is highly confident in his or her diagnosis irrespective of its conformity to or the existence of any gold standard, then any recommendation, regardless of the source, may not be influential. Overconfidence is particularly troubling because confidence is often not highly correlated with accuracy[19,20]. As a result, a physician who uses his or her confidence as a basis for disregarding advice may be making an ill-advised decision.

There are a few limitations to our study. The major limitation is the asymmetry between the description of the decision aid and the specialist's advice. Absent from written scenarios is the potential social cost of defying a colleague with whom one might have subsequent interactions. Thus, these results may not generalize to cases where the specialist is highly respected or there is a large social cost to defying the advice of the specialist.

Another potential limitation is our use of hypothetical decisions. Hypothetical decisions may not map on to real decisions; however, using a hypothetical scenario allowed us to create a case where two credible decision aids provided contradictory advice. Also, our use of residents may limit our ability to generalize these results to more experienced physicians. However, approximately one third of the residents were at the end of their final year of training, and pneumonia is a common clinical problem that most house-staff become familiar with early in their first year. This suggests that the results might also pertain to more experienced clinicians.

Future studies can attempt to tease out the salient characteristics of the decision aid or human opinion that enhance their credibility.

CONCLUSIONS

Physicians in training treating a scenario depicting community-acquired pneumonia were more influenced by the recommendation of a referenced decision aid than the recommendation of an anonymous specialist, each of which provided advice that conflicted with the initial admission decision. This suggests that greater adherence to human over decision aid advice is not a cause of decision aid non-use. Future research should further evaluate physician attitudes to decision aid use, the response to respected specialists and how justifications might influence the physician's adherence to the decision aids, as well as confirming and extending the results to other clinical scenarios and physician populations.

Acknowledgements: *We are grateful to Drs. Paneez Khoury and Kevin Patel for their help in recruiting participants and to Kenny Olson for his theoretical contribution. The project was supported by a grant from the Program in Decision, Risk, and Management Science at the National Science Foundation (NSF), grant number SES-0326468. The authors report no conflicts of interest.*

Conflict of Interest: *None disclosed.*

Corresponding Author: *Mitchell A. Medow, MD, PhD; Section of General Internal Medicine, Boston University School of Medicine, 801 Massachusetts Ave, 2nd Floor, Boston, MA 02118-2605, USA (e-mail: mitchell.medow@bmc.org).*

REFERENCES

1. **Garg AX, Adhikari NK, McDonald H, Rosas-Arellano MP, Devereaux PJ, Beyene J, Sam J, Haynes RB.** Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. JAMA. 2005;293(10):1223–38.
2. **de Dombal, Dallos V, McAdam WAF.** Can computer aided teaching packages improve clinical care in patients with acute abdominal pain? BMJ. 1991;310:1495–7.
3. **Riddderikhoff J, van Herk E.** A diagnostic support system in general practice: is it feasible? Int J Med Inform. 1997;45:133–43.
4. **Friedman DP, Elstein AS, Wolf FM.** Enhancement of clinicians' diagnostic reasoning by computer-based consultation: a multisite study of 2 systems. JAMA. 1999;282(19):1851–6.
5. **Getty DJ, Pickett RM, D'Orsi CJ, Swets JA.** Enhanced interpretation of diagnostic images. Invest Radiol. 1988;23:240–52.
6. **Chase CR, Vacek PM, Shinozaki T, Giard AM, Ashikaga T.** Medical information management: improving the transfer of research results to presurgical evaluation. Med Care. 1983;21:410–24.
7. **DesRoches CM, Campbell EG, Rao SR, Donelan K, Ferris TG, Jha A, Kaushal R, Levy DE, Rosenbaum S, Shields AE, Blumenthal D.** Electronic health records in ambulatory care—a national survey of physicians. N Engl J Med. 2008;359(1):50–60.
8. **Corey GA, Merenstein JH.** Applying the acute ischemic heart disease predictive instrument. J Fam Pract. 1987;25:127–33.
9. **Sittig DF, Krall MA, Dykstra RH, Russell A, Chin HL.** A survey of factors affecting clinician acceptance of clinical decision support. BMC Med Inform Decis Making. 2006;6:6.
10. **Arkes HR, Shaffer VA, Medow MA.** Patients derogate physicians who use a computer-assisted diagnostic aid. Med Decis Making. 2007;27:189–202.
11. **Arkes HR, Shaffer VA, Medow MA.** The influence of a physician's use of a diagnostic decision aid on the malpractice verdicts of mock jurors. Med Decis Making. 2008;28:201–8.
12. **Dreiseitl S, Binder M.** Do physicians value decision support? A look at the effect of decision support systems on physician opinion. Artif Intell Med. 2005;33(1):25–30.
13. **Fine MJ, Auble TE, Yealy DM, et al.** A prediction rule to identify low-risk patients with community-acquired pneumonia. N Engl J Med. 1997;336(4):243–50.
14. **Lim WS, van der Eerden MM, Laing R, Boersma WG, et al.** Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study. Thorax. 2003;58(5):377–82.

15. R Development Core Team (2009). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org accessed 11/25/2009.
16. **Lomas J, Enkin M, Anderson GM, et al.** Opinion leaders vs audit and feedback to implement practice guidelines. Delivery after previous cesarean section. JAMA. 1991;265(17):2202–7.
17. **Kaplan B.** Evaluating informatics applications—clinical decision support systems literature review. Int J Med Inform. 2001;64:15–37.
18. **Berner ES, Maisiak RS, Heudebert GR, Young KR Jr.** Clinician performance and prominence of diagnoses displayed by a clinical diagnostic decision support system. AMIA Annu Symp Proc. 2003;2003:76–80.
19. **Dawson NV, Connors AF Jr, Speroff T, Kemka A, Shaw P, Arkes HR.** Hemodynamic assessment in the critically ill: is physician confidence warranted? Med Decis Making. 1993;13:258–66.
20. **Landefeld CS, Chren MM, Myers A, Geller R, Robbins S, Goldman L.** Diagnostic yield of the autopsy in a university hospital and a community hospital. N Engl J Med. 1988;318:1249–54.

Effectiveness of Policies Maintaining or Restricting Days of Alcohol Sales on Excessive Alcohol Consumption and Related Harms

Jennifer Cook Middleton, PhD, Robert A. Hahn, PhD, MPH, Jennifer L. Kuzara, MA, MPH, Randy Elder, PhD, Robert Brewer, MD, PhD, Sajal Chattopadhyay, PhD, Jonathan Fielding, MD, MPH, MBA, Timothy S. Naimi, MD, MPH, Traci Toomey, PhD, Briana Lawrence, MPH, the Task Force on Community Preventive Services

Abstract: Local, state, and national laws and policies that limit the days of the week on which alcoholic beverages may be sold may be a means of reducing excessive alcohol consumption and related harms. The methods of the *Guide to Community Preventive Services* were used to synthesize scientific evidence on the effectiveness for preventing excessive alcohol consumption and related harms of laws and policies maintaining or reducing the days when alcoholic beverages may be sold. Outcomes assessed in 14 studies that met qualifying criteria were excessive alcohol consumption and alcohol-related harms, including motor vehicle injuries and deaths, violence-related and other injuries, and health conditions.

Qualifying studies assessed the effects of changes in days of sale in both on-premises settings (at which alcoholic beverages are consumed where purchased) and off-premises settings (at which alcoholic beverages may not be consumed where purchased). Eleven studies assessed the effects of adding days of sale, and three studies assessed the effects of imposing a ban on sales on a given weekend day. The evidence from these studies indicated that increasing days of sale leads to increases in excessive alcohol consumption and alcohol-related harms and that reducing the number of days that alcoholic beverages are sold generally decreases alcohol-related harms. Based on these findings, when the expansion of days of sale is being considered, laws and policies maintaining the number of days of the week that alcoholic beverages are sold at on- and off-premises outlets in local, state, and national jurisdictions are effective public health strategies for preventing excessive alcohol consumption and related harms.

(Am J Prev Med 2010;39(6):575–589) Published by Elsevier Inc. on behalf of American Journal of Preventive Medicine

Introduction

Excessive alcohol consumption in the U.S. is responsible for approximately 79,000 deaths per year, making it the third-leading cause of preventable death.[1] Approximately 15% of U.S. adults aged ≥18 years and approximately 29% of high school students in the U.S. report binge drinking (consuming five or more drinks per occasion for men, and four or more drinks per occasion for women).[2,3] The direct and indirect economic costs of excessive drinking in 1998 were $184.6 billion.[4] The reduction of excessive alcohol consumption is thus a matter of major public health and economic interest; this objective is a central goal in the U.S. public health agenda for the year 2010.[5]

This review examines the utility of enacting or maintaining limits on the days of the week on which alcoholic beverages may be sold ("days of sale") as a strategy to prevent excessive alcohol consumption and related harms. The limitation of days of sale of alcoholic beverages is here defined as "applying regulatory authority to limit the days that alcoholic beverages may be sold at on- and off-premises alcoholic beverage outlets." *Limiting* may be either maintaining existing limits (e.g., on the sale of alcoholic beverages on Sundays) or extending current limits (e.g., eliminating Sunday sales by repealing current

From the Community Guide Branch, Epidemiology and Analysis Program Office (Middleton, Hahn, Kuzara, Elder, Chattopadhyay, Lawrence), National Center for Chronic Disease Prevention and Health Promotion (Brewer, Naimi), CDC, Atlanta, Georgia; Los Angeles County Department of Public Health (Fielding), Los Angeles, California; University of Minnesota School of Public Health (Toomey), Minneapolis, Minnesota

The names and affiliations of the Task Force members are listed at www.thecommunityguide.org/about/task-force-members.html.

Address correspondence to: Robert A. Hahn, PhD, MPH, Community Guide Branch, CDC, 1600 Clifton Road, Mailstop E-69, Atlanta GA 30333. E-mail: rhahn@cdc.gov.

0749-3797/$17.00
doi: 10.1016/j.amepre.2010.09.015

authorization for such sales). Days of sale may be regulated at national, state, or local levels. *On-premises* retailing refers to the sale of alcoholic beverages for consumption at the point of sale (e.g., at bars, restaurants, or clubs); *off-premises* retailing refers to the sale (e.g., at package stores, liquor stores, grocery stores, or convenience stores) of contained alcoholic beverages for consumption elsewhere. Because most of the studies reviewed consider removing limits on days of sale (e.g., allowing sale of alcoholic beverages on Sunday when Sunday sales had previously not been allowed), the intervention of public health interest for the review is the study control condition (i.e., maintaining limits on days of sale).

In the U.S., policies restricting the days of sale currently apply to Sundays only. There are several variations on the regulation of Sunday alcohol sales in the U.S. including full bans, reduced hours relative to other days of the week, restrictions on the sale of alcoholic beverages with a high alcohol content, and the authorization of local decision making.[6] A total of fourteen states (Alabama, Arkansas, Connecticut, Georgia, Illinois, Indiana, Kansas, Kentucky, Minnesota, Nebraska, Oklahoma, South Carolina, Tennessee, and Utah), ban alcohol sales at off-premises retail alcohol outlets on Sundays. Fourteen states (Alaska, California, Colorado, Florida, Hawaii, Idaho, Kentucky, Montana, Nevada, New Hampshire, Oregon, Vermont, Wisconsin, and Wyoming) do not restrict Sunday alcohol sales. The remaining 22 states and the District of Columbia allow Sunday sales with restrictions regarding hours and/or types of alcoholic beverages sold. Outside of the U.S., current policies restricting the days of sale may apply to days other than Sunday (e.g., some countries prohibit alcohol sales on Saturdays).

In the U.S., the control of days and hours of sale at the local level is often pre-empted by state regulations prohibiting local authorities from enacting stricter alcohol control regulations in the state in general.[7,8] However, in some states, counties and other local jurisdictions are allowed to establish their own alcohol control policies. The nature of this authority varies by state and may allow cities or counties to have reduced hours from those stipulated by the state; have the same hours on Sunday as available during the rest of the week; or limit the sale of alcohol on Sundays to specific areas or locations. Fourteen states provide for local authority regarding days of sale, and four more allow Sunday sales in limited locations within the state.[6] In 1995, New Mexico repealed a ban on off-premises alcohol sales on Sundays, but also allowed local jurisdictions to hold referenda to restore a local ban on Sunday sales. Alaska and Kentucky also allow counties to independently restrict alcohol sales.

This review addresses the effects on excessive alcohol consumption and related harms of maintaining or increasing restrictions on the days of sale at on- or off-premises outlets.

Findings and Recommendations from Other Reviews and Advisory Groups

Several reviews conducted in the U.S. have concluded that restricting the days of sale is an effective strategy for reducing excessive alcohol consumption and related harms. For example, a narrative review conducted by Single[9] concluded that controlling the days (and hours) of sale may influence levels of impaired driving and other drinking problems. A systematic review published by the Substance Abuse and Mental Health Service Administration's Center for Substance Abuse Prevention[10] in 1999 found substantial evidence for harms associated with expanding the days (and hours) of alcohol sales. This finding was based on previous empirical research indicating that the expansion of the days (and hours) of sale increased prevalence of excessive alcohol consumption and alcohol-related problems. Other narrative reviews[11,12] generally concur with these findings.

Several international bodies have recommended the control of days (or hours, or both) of sale, as a means of reducing excessive alcohol consumption and related harms. The WHO has published a narrative review[13] that identifies the limiting of days of sale as an effective method for reducing alcohol-related harms. Similarly, the Western Australian Alcohol Plan[14] recommended that days and hours of sale should be considered as factors in the local regulation of alcohol availability. In Ireland, the Department of Health and Children's Strategic Task Force on Alcohol[15] concluded that "restricting any further increases in the physical availability of alcohol (number of outlets and times of sales)" is among the most effective policy measures that influence alcohol consumption and related harms.

The present review updates prior syntheses using the systematic approach of the *Guide to Community Preventive Services (Community Guide)*, as described below.

Methods

The methods of the *Community Guide* were used to systematically review scientific studies that have evaluated the effectiveness of limiting or maintaining existing limits on days of sale for preventing excessive alcohol consumption and related harms. More details on the *Community Guide* review process are presented elsewhere.[16] In brief, this process involves forming a systematic review development team; developing a conceptual approach to organizing, grouping, and selecting interventions; searching for and retrieving available research evidence on the effects of those interventions; assessing the quality of studies and abstracting information from

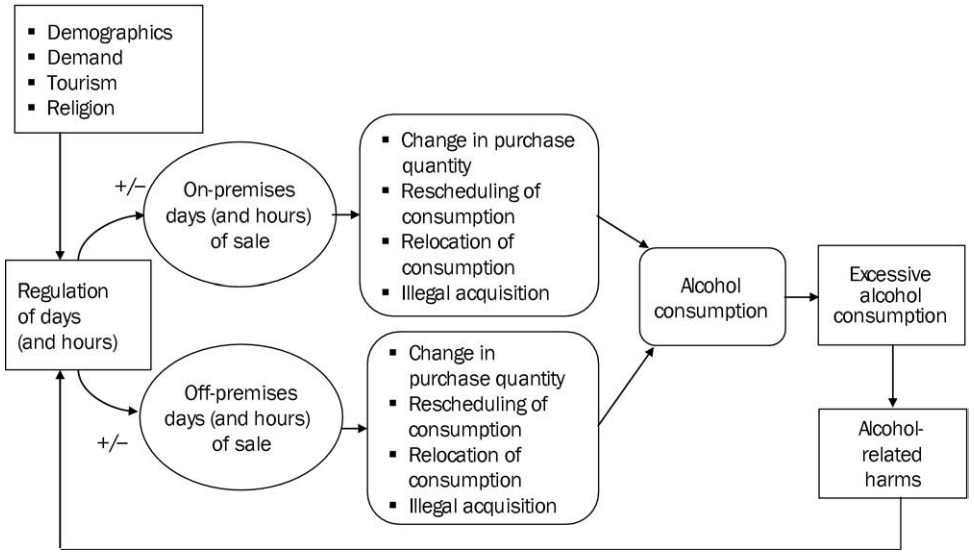


Figure 1. Effects of regulation of days (and hours) of alcohol sales on excessive alcohol consumption and related harms

each study that meets inclusion criteria; assessing the quality of and drawing conclusions about the body of evidence on intervention effectiveness; and translating the evidence on effectiveness into a recommendation or finding for each intervention reviewed. Evidence is collected and summarized on (1) the effectiveness of interventions in altering selected health-related outcomes and (2) positive or negative effects of the intervention on other health and nonhealth outcomes. To help ensure objectivity, the review process is typically led by scientists not employed by a program that might be responsible for overseeing the implementation of the reviewed intervention. When an intervention is shown to be effective, information is also analyzed on (3) the applicability of the evidence (i.e., the extent to which effectiveness data might generalize to diverse population segments and settings); (4) the economic impact of the intervention; and (5) barriers to implementation. The results of this review process are presented to the Task Force on Community Preventive Services (Task Force), a nonfederal independent scientific review board, which objectively uses specified guidelines to consider the scientific evidence on intervention effectiveness and determines whether the evidence is sufficient to warrant a recommendation.[16]

Conceptual Approach and Analytic Framework

Policies reducing or expanding days of sale (Figure 1) are hypothesized to affect alcohol consumption and alcohol-related harms through the following means: First, increases or decreases in the days of sale affect consumers' ability to purchase alcohol by changing its availability. Second, when access to alcoholic beverages changes, consumers may alter their purchasing habits in several ways, including changing their purchase volume per visit to the outlet, rescheduling their purchases, relocating their purchases, or obtaining alcoholic beverages illegally. Various characteristics of the affected population, including the demand for alcoholic beverages, the number of adult tourists the area attracts, and the religious affiliation and involvement of residents, may affect the establishment of the policies regulating days of sale.

Changes in days of sale may also affect alcohol-related outcomes by other means. For example, increases in the days of sale at on-premises outlets allow more opportunities for social aggregation, which in turn may increase aggressive behaviors that are exacerbated by alcohol consumption.[17] Increases or decreases in the days of sale may also alter travel patterns to areas where alcohol can be purchased, and thus influence the risk of injury or death in motor vehicle crashes that may be alcohol-related. It might be expected that added days of sale at on-premises outlets would be more likely to increase alcohol-related motor vehicle crashes than added days in off-premises facilities because patrons who have drunk at an on-premises facility may drive after excessive consumption, whereas patrons of off-premises outlets are not supposed to drink at that facility. It is also possible that when available days at on-premises facilities are reduced, motor vehicle crashes might be increased if consumers drove to more distant on-premises facilities and then returned after excessive consumption.

Inclusion and Exclusion Criteria

To be included as evidence in this review, studies had to

- evaluate long-term policy changes related to days of sale; studies that assessed short-term changes in alcohol availability (e.g., alcohol sales related to a special event) were not included;
- assess the impact of changes in days of sale alone on excessive alcohol consumption or related harm, as opposed to evaluating the effect of this change only in combination with other interventions;
- be conducted in a high-income country[18,a];
- present primary research findings, and not just review other research findings;
- be published in English;
- have a comparison group, or at a minimum, compare outcomes of interest before and after a change in the policy related to days of sale.

[a]Andorra, Antigua and Barbuda, Aruba, Australia, Austria, The Bahamas, Bahrain, Barbados, Belgium, Bermuda, Brunei Darussalam, Canada, Cayman Islands, Channel Islands, Cyprus, Czech Republic, Denmark, Equatorial Guinea, Estonia, Faeroe Islands, Finland, France, French Polynesia, Germany, Greece, Greenland, Guam, Hong Kong (China), Hungary, 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, Northern Mariana Islands, Norway, Oman, Portugal, Puerto Rico, Qatar, San Marino, Saudi Arabia, Singapore, Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Trinidad and Tobago, United Arab Emirates, United Kingdom, U.S., Virgin Islands (U.S.)

Table 1. Evidence of the effects of limits of days of alcohol sale on excessive alcohol consumption and related harms

Study Design description (suitability) Study execution (no. of limitations)	Population Study time period	Intervention comparison
Days of sale: On-premises		
Ligon (1996)[22] Interrupted time series: before-and-after with comparison (greatest) Fair (2)	Athens GA January 1992 –December 1993	Intervention: On 12/8/1992, Athens-Clarke County amended the Alcoholic Beverage Ordinance. Previously, Sunday sales of liquor were banned. After the change, restaurant patrons were able to purchase alcoholic beverages with food, but bars and taverns remained closed and off-premises sales were still prohibited. Comparison: Other days of the week
Smith (1978)[28] Interrupted time series: before-and-after with comparison (greatest) Fair (2)	Perth, Western Australia 3 years before and 3 years after new law (used midpoint of June 30, 1970)	Intervention: On 7/7/1970 the sale and supply of alcoholic beverages on Sundays in the Perth Metropolitan area of Western Australia became legal. In general, two 2-hour drinking sessions were permitted. Prior to the change, alcohol sales at on-premises facilities were permitted between 10 AM and 10 PM only, Monday to Saturday. Comparison: Remainder of the state
Smith (1988)[30] Before-and-after with comparison (greatest) Fair (2)	Brisbane, Australia Before period: April 1, 1968–March 31, 1970 After period: April 1, 1970–March 31, 1973 3-year After period: April 1, 1973–March 31, 1976	Intervention: On April 3, 1970, Sunday alcohol sales were introduced in Brisbane, Australia. Sunday drinking was allowed from 11 AM to 1 PM and 4 PM to 6 PM Comparison: Other days of the week and the rest of Queensland
Smith (1987)[29] Before-and-after with comparison (greatest) Fair (3)	New South Wales, Australia Before period: December 7, 1976–December 6, 1979 After period: December 7, 1979–December 6, 1981	Intervention: In 1978, Select Committee of the New South Wales Parliament considered the issue of hotel alcohol service hours in that state. Subsequently it was recommended on December 7, 1979 that the hotel service hours of 12 NOON to 10:00 PM on Sundays be introduced. Comparison: Other days of the week and the rest of the Queensland state
Smith (1990)[31] Before-and-after with comparison (greatest) Fair (3)	Victoria, Australia Before period: January 1, 1980–December 31, 1983 After period: January 1, 1984–December 31, 1984 The following 12 months were used as the "after" period for the 8-hour Sunday drinking permit.	Intervention: Two legislative changes that increased the Sunday availability of alcoholic beverages in Victoria. Prior to July 13, 1983, on Sunday, hotels and licensed clubs in Victoria could sell alcoholic beverages for consumption only with a meal. After that date, hotels and clubs were allowed to obtain a permit that permitted them to open for two 2-hour periods on Sunday between 12 NOON and 8PM. The two drinking periods had to be at least 2 hours apart. Following an amendment to the Victorian Liquor Control Act, as of November 1984, hotels and clubs could apply for a permit that enabled them to open between 12 NOON and 8 PM on Sundays. The 1984 amendment also allowed for hotels to obtain a permit to continue Monday to Saturday ordinary bar trading from 10 PM to 12 MN. The amendment also introduced Sunday restaurant hours of 12 NOON to 11:30 PM. Previously, the Sunday restaurant opening hours were 12 NOON to 4PM and 6 to 10 PM. Comparison: Other days of the week
Knight (1980)[21] Before-and-after study design without comparison (least) Fair (4)	Four major cities and central belt of Scotland Before: March 1977 After: October 1977	Intervention: In 1973, Scottish Licensing Law changed. The two main changes were the extension of evening hours on weekdays to 11 PM (previously 10 PM) and the provision for special licenses to allow pubs to open regularly on Sundays. Sunday licenses were not issued for approved public houses until October 1977. Comparison: No comparison group

(continued on next page)

Table 1. *(continued)*

Analysis Outcome	Reported findings	Review Effect size
Chi-square DUI arrests	Following the change in law, the incidence of DUIs was lowest for Sundays. The frequency of DUI arrests made on Sundays were statistically lower than every other day of the week, except for Monday	Relative % change (95%CI): 39.8 (–21.9, 150.4)
Chi-square Traffic crashes: people killed in motor vehicle crashes	Significant increase in the proportion of people killed and the number of motor vehicle crashes on Sundays, compared with the other 6 days of the week in Perth. No increases in the proportions of people killed or in the number of motor vehicle crashes occurring on Sundays in comparison with the other days of the week for the rest of the state. 11% of the 453 people killed in Perth traffic crashes were killed on Sundays: after the new law, 16.9% of 486 people were killed on Sundays (χ^2= 6.134, p>0.02). Rest of the state proportions were 18 and 17.4% before and after (χ^2= .0318, p>0.80). Motor vehicle crashes occurring on Sundays in the Perth area increased from 12.4% of 11,598 before the new law to 14.2% of 11,870 afterward (χ^2=16.85, p<0.001). In the rest of the state the proportion of motor vehicle crashes occurring on Sunday decreased from 19.7% to 18.4% (χ^2 = 15.95, p>0.20)	Relative % change: People killed: 58.9; motor vehicle crashes: 22.6
Chi-square Motor vehicle crashes	First follow-up period: Only the segment from 6:00 PM to 7:59 PM gave a significant result for Brisbane. In comparison to the other 6 days of the week, and after allowing for the slight change in the control data from the before to the after period, the annual increase was 129.8%. No significant differences in Brisbane motor vehicle crashes on Sundays between 8:00 PM and 10:59 AM. No significant increases in Queensland Sunday motor vehicle crashes occurred for any of the time segments. 3-year follow-up available, but data incomplete	Relative % change (95%CI): 65.0 (30.49, 108.65)
Chi-square Motor vehicle fatalities Traffic crashes	After the introduction of a 10-hour hotel session in New South Wales, for the 12-hour period from 12:00 NOON to 11:59 PM, there was a 22.2% increase in Sunday fatal crashes. None of the analyses for the control period of 12:00 MN to 11:59 AM gave significant results in the same direction as for motor vehicle fatalities or traffic crashes.	Relative % change (95%CI): Motor vehicle Fatalities 15.5 (–0.13, 33.59) Traffic crashes 6.7 (0.56, 13.21)
Chi-square Motor vehicle traffic crashes	The introduction of the two 2-hour drinking sessions on Sundays did not adversely affect the number of motor vehicle crashes, so information on 8-hour drinking not included.	Relative % change (95% CI): 9.9 (3.27, 16.98)
Percentage changes Consumption and patterns of consumption	Increase in consumption among men aged <45 years. Virtually no change in drinking among women.	Average change in consumption for men: 6.82 Average change in consumption for women: 1.85

(continued on next page)

Table 1. Evidence of the effects of limits of days of alcohol sale on excessive alcohol consumption and related harms *(continued)*

Study Design description (suitability) Study execution (no. of limitations)	Population Study time period	Intervention comparison
Days of sale; hours off-premises		
McMillan (2006)[23] McMillan (2007)[24] Time-series study with prospective data collection (greatest) Fair (3)	Location: New Mexico Dates: Intervention: July 1995 Pre-period: July 1990–June 1995 Follow-up: July 1995–2000	Intervention: Legalized Sunday off-premises sales: • Between the hours of 12 NOON and 12 MN • Alcohol was available on-premises prior to law change • Provision for local option to reinstate ban, municipalities to bear cost of referendum and enforcement Comparison: Pre–post study, non-Sunday days serve as control. Also comparison of alcohol- and non–alcohol-related crash trends
Norstrom (2003)[25] Norstrom (2005)[26] Experimental time-series design (greatest) Good (1)	Location: Sweden Dates: Pre-intervention: January 1995–July 2000 Phase I (experimental): February 2000–June 2001 Phase II (whole country): July 2001–July 2002	Intervention: Saturday sales allowed experimentally for six counties (Phase I) • 43% of population Saturday sales extended to whole country (Phase II) Comparison: • Seven control counties • Middle and southern regions of Sweden • 34% of population • Separated from experimental regions by buffer zone • Buffer zones 22% of population
Olsson (1982)[27] Experimental time-series design (greatest) Fair (3)	Location: Sweden Dates Pre-period: June 1980–September 1980 Follow-up: June 1981–September 1981 Intervention: May 1981	Intervention: Saturday closure of retail liquor stores Comparison: Non-Saturdays
Stehr (2007)[32] Econometric state-level time- series analysis (greatest) Fair (2)	U.S. 1990–2004 Bans were repealed in the following states: 1995: New Mexico 2002: Oregon 2003: Delaware, Kansas, Massachusetts, New York, Pennsylvania 2004: Rhode Island, Idaho, Kentucky, Ohio, Virginia 2005: Washington	Intervention: Having a Sunday ban on off-premises purchase (12 states during the study period). Specific to either beer or liquor, but wine not included. Comparison: States that did not allow sales on Sunday in each year of data collection.
Nordlund (1985)[33]	Norway Before: 1983 After: 1984	Intervention: In select villages, shops were allowed to re-open on Saturdays, in contrast to the newly instituted Saturday closing in the rest of the country. Comparison: Shops in control cities (matched by size and demographic characteristics to be similar to intervention towns). These remained open on Saturday as always.

(continued on next page)

DUI, driving under the influence

Table 1. *(continued)*

Analysis Outcome	Reported findings	Review Effect size
RR ARCs and fatalities	ARC RR (95% CI)= 1.29 (1.05, 1.58) ARC fatalities (95% CI)=1.42 (1.05, 1.93) Mean RR ARC Fatalities rest of week (95% CI)=1.07 (0.80, 1.45) Excess ARCs in study period (95%, CI)=543.1 (158.9, 927.4) Excess ARC fatalities (95% CI)=41.6 (6.6, 76.6) 2007 Study: Three counties that overturned ban repeal right away had lowest ARC RR; only one other county had RR in lowest category.	Alcohol-related Sunday crash fatalities (relative % change [95% CI]): 26.8% (3.3, 44.2)
ARIMA and parametric models Alcohol sales, assaults, drunken driving, and positive breath analyzer test results	Effects appear uniform across three experimental areas, tendency toward weaker effect in Northern Sweden not sign. Phase I consumption (relative % change): ● Beer (7.6%) ● Wine (2.5%) ● Spirits (3.7%) ● Total alcohol (3.7%) Phase II consumption (relative % change): ● Beer (1.8%) ● Wine (1.2%) ● Total alcohol (3.6%)	Relative % changes (95% CI): Drunk driving: 11.3% (4.2, 18.4) Alcohol sales (liters pure alcohol per capita per year): 3.6% (2.6, 4.6) Assaults, women (indoors): 0.6% (–6.5, 7.7) Assaults, total: –1.3% (–5.6, 3.0)
Police interventions Intoxicated people Domestic disturbances Outdoor assaults	Sales of alcohol: Slight decline could not be attributed to effects of Saturday closing. Illegal trading: (Police judgment) % of districts reporting: ● No change: 69% ● Increase: 24% ● Decrease: 7% Overall declines in: ● Drunkenness ● Domestic disturbances ● Public disturbances (not attributable to policy) ● Assaults declined	Relative % changes (95% CI): Outdoor assaults: –17.7% (−45.8, 25.0) Domestic disturbances –17.3% (–34.8, 4.8) Police interventions against intoxicated people –35.7% (–43.8, –26.4)
Time-series analysis of state-level variables, including Sunday bans. Controlled for pre-repeal trends	Per capita beer sales in gallons ● –2.4 relative % change due to Sunday bans controlling for pre-repeal trends ● –4.1 relative % change due to Sunday ban not controlling for pre-repeal trends Per capita spirits sales in gallons ● –3.5 relative % change due to Sunday ● –5.2 relative % change due to Sunday ban not controlling for pre-repeal trends.	Beer sales: 2.4% relative change due to repeal of bans Spirits sales: 3.5% relative change due to repeal of bans Note: Although authors coded for presence of Sunday bans, all policy changes during the study period were in the direction of repeal, so the signs have been reversed in reporting effect (above).
Customer calls Cash turnover Liters pure alcohol Liters total sale all outlets Arrests for drunkenness Reports of drunkenness Reports domestic trouble Reports of violence	Customers made fewer trips to vinmonopolets (i.e., state alcoholic beverage monopoly stores). Total sales at these outlets declined, but the total sales at all outlets went up slightly. Reports of drunkenness went down but not significantly, while drunkenness arrests declined significantly. Reports of domestic trouble went down a sizeable and significant 16%, whereas reports of violence went up 5%. General effects were consistent but small; ordinary drinkers consumed about the same total amount, purchased in fewer trips to the vinmonopolets with larger purchases per trip. Ultimately, the Saturday closing was repealed because of insufficient evidence of benefit.	Relative % changes: Liters pure alcohol: –3.1% Arrests for drunkenness: –5.8% Reports of drunkenness: –5.0% Reports domestic trouble: –15.9% Reports of violence: 5%

ARC, alchohol-related crashes; ARIMA, autoregressive integrated moving average; RR, relative risk

To be included in this review, studies also had to report on outcomes related to excessive alcohol consumption or related harms. Specific types of harm that were of interest included alcohol-related medical conditions (e.g., liver cirrhosis); alcohol-impaired driving; alcohol-related crashes; unintentional or intentional injuries; and violent crime.

Outcome measures that had the strongest known association with excessive alcohol consumption included binge drinking, heavy drinking, liver cirrhosis mortality, alcohol-related medical admissions, and alcohol-related motor vehicle crashes, including single-vehicle night-time crashes (which are widely used to indicate motor vehicle crashes due to drinking and driving).[19] Less-direct measures included per capita alcohol consumption, a recognized proxy for estimating the number of heavy drinkers in a population[13,20]; unintentional injuries; suicide; and crime, such as homicide and aggravated assault. When studies assessed multiple outcomes of interest, those outcomes with the strongest known association with excessive alcohol consumption were selected.

Search for Evidence

The following databases were searched from inception to February 2008: Econlit, PsycINFO, Sociology Abstracts, MEDLINE, Embase, and EtOH. Searches also were conducted of the reference lists of papers reviewed as well as lists in review articles. Government reports were considered for review, but other unpublished papers were not. In addition, experts were consulted to identify other studies that might have been missed.

Assessing the Quality and Summarizing the Body of Evidence on Effectiveness

Each study that met the inclusion criteria was read by two reviewers who used standardized criteria (available at 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 the team members.

Studies were evaluated based on their design and execution. The current classification of the study designs accords with *Community Guide* standards[16] and may differ from the classification reported in the original studies. Those that collected data prospectively on exposed and control populations were classified as having the greatest design suitability. Those that collected data retrospectively or lacked a comparison group but that conducted multiple pre- and post-measurements on their study population(s) were rated as having moderate design suitability. Finally, cross-sectional studies, those without a comparison group, and those that involved only a single pre- and post-measurement in the intervention population were considered to have the least suitable design. Quality of execution was assessed by examining potential threats to study validity, including an inadequate description of the intervention or of the study population, poor measurement of the exposure or outcome, failure to control for potential confounders, and a high level of attrition among study participants. Based on these criteria, studies were characterized as having good quality of execution if they had at most one threat to validity, fair execution if they had two to four threats to validity, and limited quality of execution if they had five or more threats to validity. Only studies with good or fair quality of execution were included in the body of evidence; studies with any level of design suitability were included, other than those with cross-sectional design.

We calculated effect sizes as relative percentage change in the intervention population compared with the control population using the following formulas:

- For studies with before-and-after measurements and concurrent comparison groups:

  $$\text{Effect size} = [(I_{post}/C_{post})/(I_{pre}/C_{pre}) - 1] \times 100\%,$$

  where:

  I_{post} = last reported outcome in the intervention group after the intervention;

  I_{pre} = reported outcome in the intervention group before the intervention;

  C_{post} = last reported outcome in the comparison group after the intervention;

  C_{pre} = reported outcome in the comparison group before the intervention.

- For studies with before-and-after measurements but no concurrent comparison:

  $$\text{Effect size} = [(I_{post} - I_{pre})/I_{pre}] \times 100\%$$

  When there was a large enough number of studies of a single outcome, median effect size and interquartile intervals were reported.

Results

Intervention Effectiveness

Fourteen studies[21–34] that examined the effects of changes in days of sale met the inclusion criteria for the review. These studies assessed changes that took place in cities (Athens GA [two studies] and Perth and Brisbane, Australia); states (50 U.S. states, New Mexico [two studies], and Victoria and New South Wales, Australia); and countries or large regions of countries (Norway [one study], Sweden [three studies], and Scotland [one study]). The policy changes that were assessed took place between 1967 and 2004. (For a summary of all evidence included in this review, see Table 1.)

The studies used a variety of methods for estimating intervention effects, including chi-square statistics, percentage change, relative risks, and auto-regressive integrated moving average (ARIMA) time series; all except one study[21] had comparison populations or conditions. Thirteen studies[22–34] were of greatest design suitability and one[21] was of least design suitability. Four studies[25,26,32,33] were of good execution and the remainder[21–24,27–31,34] were of fair execution. Studies assessing changes in days of sale in off-premises settings were analyzed separately from those in on-premises settings. Four studies[28–31] were conducted by one researcher (Smith), and two studies each by Ligon and Thyer,[22,34] McMillan and colleagues,[23,24] and Norstrom and Skog.[25,26]

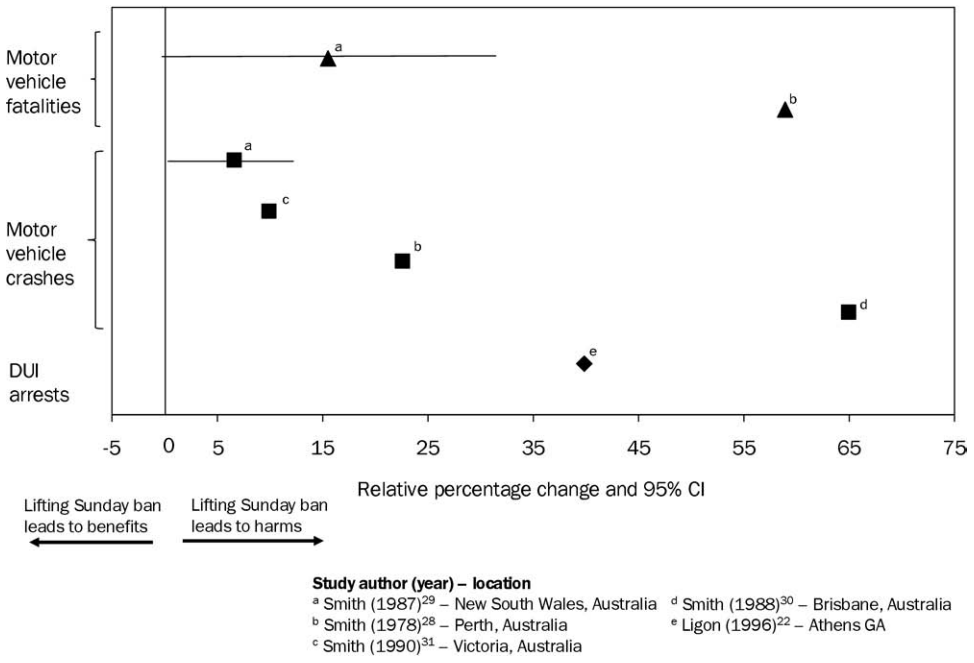


Figure 2. Relative percentage change in motor vehicle–related events after Sunday on-premises sales legalized
DUI, driving under the influence

The Effect of Changing the Number of Days That Alcohol Was Sold at On-Premises Outlets

Seven studies[21,22,28–31,34] assessed the effects of increasing days of sale at on-premises retail alcohol outlets. Only one study[21] assessed changes in consumption; the remainder assessed the effects of changes in days of sale on motor vehicle–related outcomes.

Effect on excessive alcohol consumption. The findings of Knight and Wilson[21] were reviewed in detail because only these authors examined excessive consumption among individuals (rather than per capita consumption or alcohol-related harms). This study assessed the impact on excessive alcohol consumption of a 1977 law allowing Sunday alcohol sales in the four major cities and within the central belt of Scotland. After Sunday pub sales were legalized in this area, there was a 1.3 (95% CI=−0.4, 2.8) standard unit of alcohol (a British measure equivalent to 0.6 of the U.S. standard drink) increase in the average weekly consumption by men who drank; a significant 2.4 standard unit (95% CI=0.6, 4.2) increase among men aged 18–45 years; and a nonsignificant −0.5 (95% CI=−2.6, 1.3) standard unit change in the average weekly consumption of men aged >45 years. Increases among men occurred across most levels of baseline drinking. The researchers reported a nonsignificant −0.6 standard unit change among women who drank (95% CI=−1.6, 0.5) that did not differ by age. Knight and Wilson also obtained information on the patterns of consumption among study participants. After the change, the percentage of people who reported having 1–8 standard units on Sundays increased from 27% to 29% (7.4%, 95% CI=−11.0, 31.1), and those who reported having >8 standard units increased from 4% to 5% (25%, 95% CI=−26.5, 100.1); neither increase was significant.

Effect on alcohol-impaired driving and motor vehicle crashes. Five studies[22,28–31] examined the impact of allowing Sunday on-premises sales on various measures of alcohol-impaired driving (e.g., arrests for driving under the influence [DUI]) and motor vehicle crashes [Figure 2]). An additional study in Athens GA[34] examined the impact of a December 1992 local law that allowed Sunday sales in restaurants (but not in bars). The investigators found that this change was followed by a 39.8% increase in DUI arrests (95% CI not calculable).

Two studies[28,30] assessed the impact of changes in days of sale in on-premises retail outlets in Perth and Brisbane, Australia, on deaths and injuries related to motor vehicle crashes; they compared outcomes on days when alcohol became newly available with outcomes on days when availability did not change. The city of Perth legalized Sunday alcohol sales in 1970, allowing two 2-hour periods when alcoholic drinks could be purchased. After this change, there was a 22.6% increase in motor vehicle crashes and a 58.9% increase in motor vehicle fatalities in Perth compared with the rest of the state. In the same year, Sunday sales were legalized in Brisbane also, resulting in an increase of 65% (95% CI not calculable) in motor vehicle crashes.

Finally, two additional studies assessed the effects on motor vehicle crashes of allowing Sunday sales in different regions of Australia. In 1979, the state of New South Wales began allowing hotels to serve alcoholic beverages between 12 NOON and 10 PM on Sundays.[29] This change was followed by an increase of 6.7% (95% CI=0.6%, 13.2%) in traffic crashes and an increase of 15.5% (95% CI=−0.1%, 33.6%)

in motor vehicle fatalities, compared with other days of the week in which hours did not change. Lastly, a study by Smith[31] assessed the influence of newly legalized Sunday sales in clubs and hotels on motor vehicle injury crashes in the state of Victoria. Before the law changed in 1983, hotels and licensed clubs could sell alcoholic beverages only with a meal. After the law changed, a meal was no longer required for the consumption of alcohol, and two 2-hour drinking periods were introduced. In the following year, there was a 9.9% increase in motor vehicle crashes on Sundays compared with days of the week in which hours had not changed (95% CI=3.3%, 17.0%).

Effect of Changing the Number of Days That Alcohol Was Sold at Off-Premises Outlets

Effect of repealing bans on days of sale. Four studies[23,25,26,32] examined the impact of increasing the days of sale at off-premises locations (Figure 3), by removing existing restrictions. Two of these studies[25,26] examined the two-phase reinstatement of Saturday sales in Sweden between 2000 and 2003 (Sunday sales remained banned). Another study[23] examined the repeal of a ban on Sunday sales in New Mexico. Lastly, a time-series study[32] examined the impact of bans across U.S. states over a period of 15 years, during which policies on off-premises Sunday sales changed in 13 states.

One study[25] examined the effect of removing a nearly 20-year ban on Saturday alcohol sales at off-premises locations in Sweden. Researchers collaborated with the Swedish government to implement a national experiment. In the first phase, to assess possible harms, Saturday sales were allowed only in select counties for an experimental period of 1 year. The intention was to repeal the ban on Saturday sales in the rest of the country if harms did not increase significantly when the repeal was in place in the experimental counties. To limit confounding by cross-border sales, buffer zones were designated between the experimental areas and the control areas. The experimental areas were noncontiguous, and included several rural areas, as well as Stockholm, encompassing about 43% of the population. The control area covered seven contiguous counties and another eight counties not contiguous with those, with a total of about 34% of the population. The buffer counties had approximately 22% of the population.

During Phase I, alcohol sales in the experimental area increased 3.6% (95% CI=2.6%, 4.6%) and incidents of drunk driving arrests increased by 11.3% (95% CI=4.2%, 18.4%) compared with that in the control areas. Both findings were significant. However, the researchers noted that along with repeal of the ban, there was increased police surveillance for alcohol-related motor vehicle incidents in the experimental region, which may have contributed to the increase in the number of drunk driving incidents reported. Assaults against women indoors (a proxy for domestic violence) increased 0.6% (95% CI=−6.5%, 7.7%) and total assaults declined by 1.3% (95% CI=−5.6%, 3.0%); neither result was significant.

During Phase II, the repeal of the ban on Saturday sales was extended to the whole country.[26] Alcohol sales increased by 3.5% (95% CI=3.0%, 4.0%) in what had been the control and buffer regions in Phase I—an increase similar to that which had occurred in experimental counties in Phase I. The 1.7% (95% CI=−7.0%, 10.0%) in-

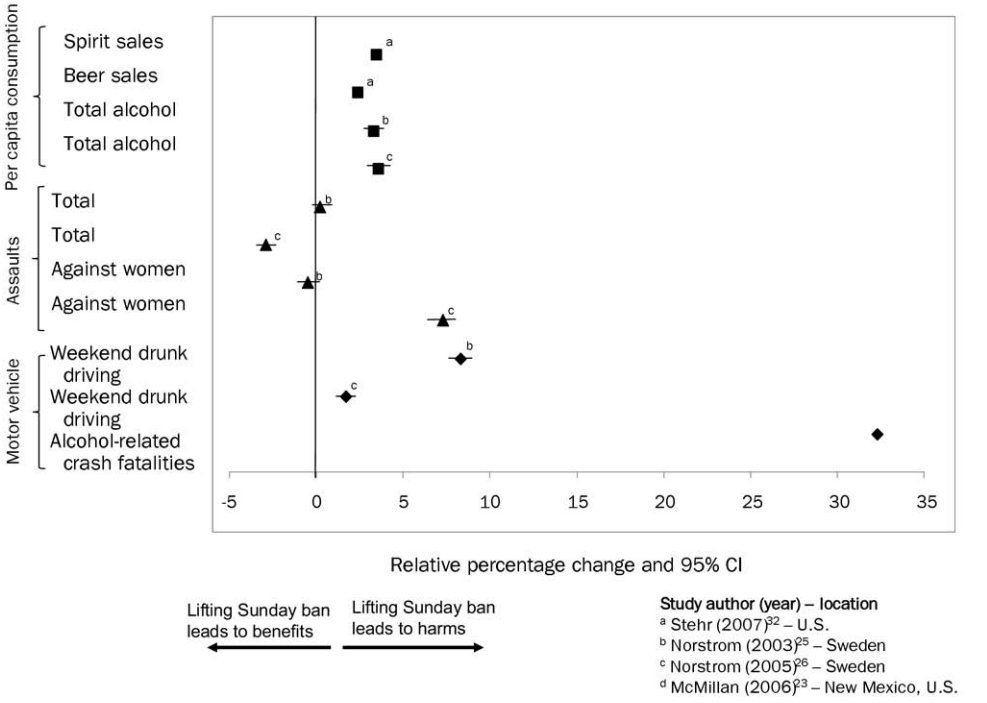


Figure 3. Relative percentage change in three categories of alcohol-linked effects attributable to an increase in days of alcohol sale in off-premises establishments

crease in drunk driving arrests in the rest of the country was not significant in Phase II (unlike in Phase I).

McMillan and others[23] examined the impact of the repeal of a ban on Sunday alcohol sales at off-premises retail outlets in New Mexico in 1995. (On-premises consumption of alcohol on Sundays was allowed already in New Mexico at that time, and was not changed by the law.) The study evaluated the impact of this change on deaths in alcohol-related motor vehicle crashes. Crashes were considered to be alcohol-related if one of the drivers involved in the crash had a blood alcohol concentration (BAC) >0.0%. To assess the impact of the repeal on alcohol-related crash fatalities, the researchers calculated the relative risk of dying in an alcohol-related crash, by day of the week, after alcohol sales were allowed on Sundays compared with the period prior to the change. They then compared the relative risk of death in an alcohol-related crash on Sundays (RR=1.4) to the mean relative risk of death in an alcohol-related crash on other days of the week (RR=1.1). Thus, the risk of death in an alcohol-related crash on Sunday increased 26.8% (95% CI=3.3%, 44.2%) relative to the risk of death in a crash on other days of the week after the ban on Sunday alcohol sales was repealed.

Finally, one study[32] examined state-level U.S. data to determine the impact on beer and liquor consumption of laws repealing bans on Sunday alcohol sales in states. The authors used a time-series analysis to compare changes from 1990 to 2004 in per capita alcohol consumption in 13 states that repealed bans on Sunday alcohol sales relative to changes in consumption in other states that maintained existing state policies on Sunday sales. Controlling for other variables such as income and taxes, as well as trends in alcohol consumption in the 13 states before the bans were repealed, the researchers found that per capita spirits consumption was 3.5% higher in states that allowed Sunday sales of spirits than in states that did not. In six states that allowed Sunday sales of beer, beer consumption was 2.4% higher.

Effects of imposing bans on days of sale. Three studies[24,27,33] examined the effect of imposing bans on days of sale of alcoholic beverages for off-premises purchase. One of these[27] examined the impact of the 1981 imposition of the Saturday ban on off-premises alcohol sales in Sweden that was discussed above. A second examined the impact of the 1984 imposition of a Saturday ban on alcohol sales in Norway.[33] The third examined the local referendum-based re-imposition of a previously repealed state ban on Sunday sales, described above, in several New Mexico counties.[24]

Olsson and colleagues[27] compared outdoor assaults, domestic disturbances, and police interventions against intoxicated people during the ban with the same 3-month period in the previous year when the ban was not in place. They also compared the number of these events that took place on Saturdays with the number of events that took place during the rest of the week over these two 3-month periods. During the ban, outdoor assaults on Saturdays declined by 17.7% (95% CI=−25.7%, –8.9%) relative to the rest of the week from a mean of 71.0 assaults per Saturday in the nation before the policy change to 53.2 after, compared with a mean change from 27.8 to 25.3 for the rest of the week. Domestic disturbances similarly declined by 17.3% (95% CI=−22.0%, –12.4%) relative to the rest of the week from a mean of 205.6 domestic disturbances per Saturday prior to the policy change to 154.9 per Saturday after, compared with a mean change of 104.5 to 95.3 for the rest of the week. During the ban, police interventions against intoxicated people declined by 35.7% (95% CI=−37.8%, –33.5%) relative to the rest of the week from 659.8 per Saturday before to 401.1 per Saturday after the policy change, compared with a mean change of 453.6 to 428.8 for the rest of the week.

In 1984, the Norwegian government initiated a similar experimental ban to determine whether closing state-run spirits and wine monopoly stores on Saturdays would reduce alcohol-related harms.[33] Because it was available from other sources, beer remained available on Saturdays during the experimental period. Six pairs of Norwegian communities in similar settings and with similar demographics were selected, with one community in each pair randomly selected for the intervention, and the other for the control. Nordlund evaluated changes in consumption and alcohol-related harms in October 1984, before completion of the experimental intervention year. Compared with the control communities, the consumption of ethanol (from wine and spirits) decreased by 3.1% in the experimental communities. However, the consumption of beer increased by a relative 6.4%, for a combined relative increase of total alcohol consumption of 0.7% in the experimental settings. In addition, there were relative declines of 5.8% in arrests for drunkenness and 15.9% in domestic trouble, but a relative increase of 5.0% in reports of violence in experimental communities compared with control communities. In sum, there was little net change in alcohol consumption associated with the ban and mixed results in terms of other alcohol-related outcomes. The Norwegian government concluded that the closing had little substantial effect and reverted to the prior policy allowing Saturday retail sales.

Finally, in addition to their analysis of repeal of the New Mexico ban on Sunday alcohol sales, described above, McMillan and colleagues undertook an analysis of data on the effects of local reinstatement of the ban.[24] The 1995 New Mexico law allowed local communities to re-

instate the Sunday sales ban following a community referendum (mounted at community expense). The towns of Gallup, Clovis, and Portales reinstated the ban within 3 months after the statewide repeal. Each of these cities is the county seat, and each comprises a sizable proportion of the total county population (70%, 27%, and 62%, respectively), such that county-level data can be taken as a gross measure of the impact of the local decision passed by these cities. Each of the three counties that rapidly reversed the state policy locally had a relative risk of Sunday alcohol-related motor vehicle crashes (comparing crash levels in each county after the policy change to levels before the change) between 1 and 1.13, the lowest reported relative risks among counties in the state. Of 33 total counties in New Mexico, only one other county had a relative risk in that range. Three other towns passed local policies somewhat later. One, Roswell, which makes up 74% of its home county, had a relative risk of <1.30. The remaining two towns had populations <2000, and would therefore not be expected to show a stable effect at the county level.

In sum, the findings from these three studies indicate that local decisions to reinstate a 1-day off-premises sales ban protected against the alcohol-related harms observed in areas that maintained the state (no ban) policy. The researchers note that these findings were based on a small number of communities and few years of data.

Conclusion

This review found that increasing days of sale by allowing previously banned alcohol sales on either Saturdays or Sundays increased excessive alcohol consumption and related harms, including motor vehicle crashes, incidents of DUI, police interventions against intoxicated people, and, in some cases, assaults and domestic disturbances. Thus, *maintaining* existing limits on Saturday or Sunday sales—the control condition in these studies—can prevent alcohol-related harms that would be associated with increased days of sale. A study of the imposition of a Saturday ban in Norway showed mixed effects, whereas one study of the imposition of a Saturday ban in Sweden and one study of the reversal of a lifted ban in New Mexico found a decrease in alcohol-related harms. Thus, some evidence suggests that *imposing* limits on the days of sale will reduce alcohol-related harms.

According to the *Community Guide* rules of evidence, there is strong evidence for the effectiveness of maintaining limits on days of sale for the reduction of alcohol-related harms. Of the qualifying studies on the *repeal of weekend-day sale bans* evaluated by *Community Guide* criteria, there were nine of greatest design suitability, three of which were of good execution and six of fair execution; there was one study of least-suitable design and fair execution. Most findings in this body of evidence indicated harms associated with an increased day of sale; effect sizes were of public health significance.

There were three studies of greatest design suitability and fair execution that assessed the impact of *imposing bans* on weekend days of sale. Two of these studies indicated that restricting days of sale is associated with a decrease in excessive alcohol consumption and related harms, and the third did not. By *Community Guide* standards, there is not sufficient evidence on which to base a determination of effectiveness. However, these studies support the overall conclusion that increasing days of sale is directly associated with excessive alcohol consumption and related harms.

Other Harms and Benefits

In association with fewer days of sale and reduced consumption, community quality of life—evaluated through such factors as reduced levels of public drunkenness—may improve on days when alcohol outlets are closed. Although it is possible that crimes such as illicit alcohol production and sales may increase in localities in which days of sale are reduced, no evidence of such effects was found.

Applicability

The studies in this review were conducted in a variety of settings in the U.S. and in other countries and during a wide range of time periods. The association between restrictions on days of sale and excessive alcohol consumption and related harm was consistent across most geographic locations and time periods. Moreover, three of the studies of greatest design suitability were conducted in the U.S. and were published within the past 10 years. Thus, the findings of this review are relevant for examining the potential impact of current proposals to modify days of sale in the U.S.

Barriers

Reductions in days of sale and resulting reductions in excessive alcohol consumption and related harms may affect overall alcohol sales; thus those restrictions may be opposed by firms involved in manufacturing, distributing, or selling alcoholic beverages. Indeed, the alcohol industry has tended to support policies removing restrictions on days of sale,[35] although some industry groups or individual businesses have supported the maintenance of Sunday sales bans.[36]

State pre-emption laws (i.e., laws that prevent the implementation and enforcement of more restrictive local alcohol sales laws) can also undermine efforts by local governments to regulate days of sale.[7] The elimination of

pre-emption laws related to the sale of tobacco products was one of the health promotion objectives in *Healthy People 2010*[5]; however, *Healthy People 2010* had no similar objective related to eliminating pre-emption of the local regulation of alcohol sales.

Economics

We identified one study[37] that assessed the economic impact of reducing days of sale. This study modeled the cost effectiveness of restricting alcohol sales for a 24-hour period over the weekend in 12 global health regions, as defined by the WHO. The costs associated with this intervention included the cost of passing the legislation itself, and the cost of administering and enforcing the laws once passed. Effectiveness was assessed using Disability-Adjusted Life Years (DALYs), a standard measure of global health impact that considers the impact of an intervention on healthy years of life lost due to either death or disability. For the region most relevant to this review, the America's A region composed of the U.S., Canada, and Cuba, the estimated cost for limiting weekend days of sale was $175,616 (converted to 2007 dollars using the Consumer Price Index) per 1 million population per year, based on a 10-year implementation period and discounted at 3%. At the same time, this restriction was estimated to prevent the loss of 250 DALYs per 1 million population per year, yielding an average cost-effectiveness ratio for this intervention of approximately $700 per DALY averted, which is much less than the average annual income per capita in these three countries, a threshold for an intervention to be considered very cost effective that was proposed by the Commission on Macroeconomics and Health.[38] To obtain country-specific estimates of the DALYs saved per country as a result of this intervention, the regional analysis needs to be adjusted using country-specific data. Such estimates are limited by data available and based in part on assumptions made.

We found no study that specifically estimated the magnitude of commercial losses in sales and tax revenues resulting from a policy of restricting days of sale. Regarding the economic burden of such a policy in terms of premature mortality, the one study that examined the impact of lifting a Sunday packaged alcohol sales ban in New Mexico[23,24] showed that this policy resulted in an estimated increase of 41.6 alcohol-related fatalities on Sundays for the 5-year period from 1995 to 2000, which translated to more than $6 million of additional cost per year for the state when the team applied the approximate unit cost of $745,285 (in 2007 dollars)[39] per motor vehicle fatality.

Research Gaps

The research on days of sale conducted in the U.S. was primarily at the state level. However, additional research is needed to assess the effectiveness of local restrictions on days of sale in preventing excessive alcohol consumption and related harms.

It would be useful to better understand the effect of differential policies regarding days of sale across neighboring jurisdictions. Does more ready access in a neighboring region lead to increased travel to this region, allowing the possibility of motor vehicle crashes, especially with intoxicated drivers?

Additional research is also needed to more fully assess the costs and benefits of restricting the number of days of sale. From a societal perspective, these should include intervention costs; loss in sales and tax revenues and employment; reductions in fatal and nonfatal injuries, crime, and violence; gains in safety and public order; and averted loss of household and workplace productivity.

Discussion

We found strong and consistent evidence that limiting alcohol availability by maintaining existing limits on the days of sale is an effective strategy for preventing excessive alcohol consumption and related harms. In addition, there is some direct evidence that the imposition of increased limits on days of sale may reduce alcohol-related harms. However, further scientific evidence is needed to fully assess the symmetry between maintaining existing limits and implementing new restrictions on days of sale, specifically as regards the impact of the latter on excessive alcohol consumption and related harms.[b]

In addition to the small number of studies that assessed the effect of new restrictions on days of sale, the studies in this review had several other limitations. First, some studies did not directly assess the impact of restrictions on days of sale on excessive alcohol consumption and related harms, but rather relied on proxy measures of these out-

[b]A reviewer of this manuscript indicated two studies of the effects of expanding days of sale published after the close of our reference search in February 2008: Carpenter 2009 and Stehr 2010. (Carpenter CS, Eisenberg D. Effects of Sunday sales restrictions on overall and day-specific alcohol consumption: evidence from Canada. J Stud Alcohol Drugs 2009;70(1): 126—33; Stehr M. The effect of Sunday sales of alcohol on highway crash fatalities. B.E. Journal of Economic Analysis & Policy 2010;10.1.) Both studies assess the effects of expanded days of sale in off-premises facilities, for which we hypothesize smaller effects. In a cross-sectional study, Carpenter finds increased consumption on Sundays in Canadian provinces with newly allowed Sunday sales, compared with provinces which maintain Sunday sales prohibition; however, there are also reductions in consumption on other days, yielding no net effect. Stehr, who in an earlier study included in our review indicated increased consumption associated with newly allowed Sunday sales in U.S. states, in this recent study finds increases in automobile crashes in New Mexico, but not in other states. These recent studies are not entirely consistent with earlier research and suggest a need for additional research.

comes (e.g., motor vehicle crashes not specifically related to alcohol). In these cases, focus was placed on measures for which the links between proxy and health outcome have been well established. Second, these studies were often unable to control for some potential confounding factors. However, they generally assessed changes in the same geographic area and within a fairly short time period before and after the implementation of changes in days of sale. Consequently, other contextual factors that could influence alcohol sales and consumption (e.g., changes in alcohol excise taxes) at the country, state, or community levels were likely to have remained fairly constant during the study periods, thus allowing for a more valid assessment of the impact of changing days of sale on excessive alcohol consumption and related harms.

One issue not addressed in this review is the potential consequence of neighboring regions having differing policies. For example, if one community restricts access to alcohol by not allowing sales on certain days, although the neighboring community lacks these restrictions, it is possible that harms (e.g., crashes from driving, drunk or sober, over longer distances) may result when those in the restricted neighborhood travel to the other community.

The findings in this review also support the potential value of allowing local communities to maintain restrictions on days of sale independent of state policies preemptively regulating days of sale. If further research supports the effectiveness of local restrictions on days of sale, it would also argue for eliminating state pre-emption laws that prohibit local governments from enacting alcohol control policies that are more restrictive than those that exist statewide.

We acknowledge the support and contributions of Steven Wing (Substance Abuse and Mental Health Services Administration).

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 Mortal Wkly Rep 2004;53(37):866–70.
2. National Center for Health Statistics. Health, U.S., 2005 with chartbook on trends in the health of America. 2005. www.cdc.gov/nchs/data/hus/hus05.pdf.
3. 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.
4. 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. Rockville MD: NIAAA, 2000. Report No.: 98-4327.
5. USDHHS. Healthy people 2010. www.healthypeople.gov/. 2001.
6. DISCUS (Distilled Spirits Council of the U.S.). Sunday spirits sales: Rolling back the Blue Laws. www.discus.org/issues/sunday.asp. 2008.
7. Mosher J. Alcohol issues policy briefing paper: the perils of preemption. Chicago: American Medical Association, 2001.
8. Gorovitz E, Mosher J, Pertschuk M. Preemption or prevention? Lessons from efforts to control firearms, alcohol, and tobacco. J Public Health Policy 1998;19(1):36–50.
9. Single E. Public drinking, problems and prevention measures in twelve countries: results of the WHO project on public drinking. Contemp Drug Prob 1997;24:425–48.
10. Division of State and Community Systems Development, Center for Substance Abuse Prevention. Preventing problems related to alcohol availability: environmental approaches. Rockville MD: Substance Abuse and Mental Health Services Administration, 2008. DHHS Publication No.: (SMA)99–3298.
11. Stockwell T, Gruenewald P. Controls on the physical availability of alcohol. In: Stockwell T, Heather N, eds. The essential handbook of treatment and prevention of alcohol problems. New York: John Wiley & Sons, 2004:213–33.
12. Smith DI. Effectiveness of restrictions on availability as a means of preventing alcohol-related problems. Contemp Drug Prob 1988; 627–84.
13. Babor TF, Caetano R, Casswell S, et al. Alcohol: no ordinary commodity-research and public policy, 2nd edition. Oxford, UK: Oxford University Press, 2010.
14. WHO. European Alcohol Action Plan, 2000–2005. Copenhagen, Denmark: WHO Regional Office for Europe, 2000.
15. Strategic Task Force on Alcohol. Strategic Task Force on Alcohol—second report. Ireland: Health Promotion Unit, Department of Health and Children, 2004.
16. Briss PA, Zaza S, Pappaioanou M, et al. Developing an evidence-based Guide to Community Preventive Services—methods. Am J Prev Med 2000;18(1):35–43.
17. Lipsey MW, Wilson DB, Cohen MA, Derzon JH. Is there a causal relationship between alcohol use and violence? In: Galanter M, ed. Recent developments in alcoholism: volume 13, alcohol and violence. New York: Plenum Press, 1997:245–82.
18. World Bank. World Development Indicators 2006. devdata.worldbank.org/wdi2006/contents/cover.htm. 2006.
19. Gruenewald PJ, Millar AB, Treno AJ, Yang Z, Ponicki WR, Roeper P. The geography of availability and driving after drinking. Addiction 1996;91(7):967–83.
20. Cook PJ, Skog OJ. Alcool, alcoolisme, alcoolisation—comment. Alcohol Health Res World 1995;19(1):30–1.
21. Knight I, Wilson P. Scottish licensing laws. London: Office of Population Censuses and Surveys, Social Survey Division, 1980.
22. Ligon J, Thyer BA, Lund R. Drinking, eating, and driving: evaluating the effects of partially removing a Sunday liquor sales ban. J Alcohol Drug Educ 1996;42(1):15–24.
23. McMillan GP, Lapham SC. Effectiveness of bans and laws in reducing traffic deaths: legalized Sunday packaged alcohol sales and alcohol-related traffic crashes and crash fatalities in New Mexico. Am J Public Health 2006;96(11):1944–8.
24. McMillan GP, Hanson TE, Lapham SC. Geographic variability in alcohol-related crashes in response to legalized Sunday packaged alcohol sales in New Mexico. Accid Anal Prev 2007;39(2):252–7.
25. Norstrom T, Skog OJ. Saturday opening of alcohol retail shops in Sweden: an impact analysis. J Stud Alcohol 2003;64(3):393–401.
26. Norstrom T, Skog OJ. Saturday opening of alcohol retail shops in Sweden: an experiment in two phases. Addiction 2005;100(6):767–76.
27. Olsson O, Wikstrom PH. Effects of the experimental Saturday closing of liquor retail stores in Sweden. Contemp Drug Prob 1982;325–53.

28. Smith DI. Impact on traffic safety of the introduction of Sunday alcohol sales in Perth, Western Australia. J Stud Alcohol 1978;39(7):1302–4.
29. Smith DI. Effect on traffic accidents of introducing Sunday hotel sales in New South Wales, Australia. Contemp Drug Prob 1987;279–94.
30. Smith DI. Effect on traffic accidents of introducing Sunday alcohol sales in Brisbane, Australia. Int J Addict 1988;23(10):1091–9.
31. Smith DI. Effect on casualty traffic accidents of changing Sunday alcohol sales legislation in Victoria, Australia. J Drug Issues 1990; 20(3):417–26.
32. Stehr M. The effect of Sunday sales bans and excise taxes on drinking and cross-border shopping for alcoholic beverages. Natl Tax J 2007;60(1):85–105.
33. Nordlund S. Effects of Saturday closing of wine and spirits shops in Norway. Oslo, Norway: Statens institutt for alkoholforskning, 1985.
34. Ligon J, Thyer B. The effects of a Sunday liquor sales ban on DUI arrests. J Alcohol Drug Educ 1993;38(2):33–40.
35. Giesbrecht N. Roles of commercial interests in alcohol policies: recent developments in North America. Addiction 2000;95(4):S581–95.
36. Hoover T. Sunday alcohol sales set—Ritter signs the bill over objections from grocers and convenience stores. The Denver Post 2008, April 15.
37. Chisholm D, Rehm J, Ommeren MV, Monteiro M. Reducing the global burden of hazardous alcohol use: a comparative cost-effectiveness analysis. J Stud Alcohol 2004;65:782–93.
38. Sachs JDC. Macroeconomics and health: investing in health for economic development. Report of the Commission on Macroeconomics and Health. Geneva: WHO, 2001.
39. CDC. Economic impact of motor-vehicle crashes—U.S., 1990. MMWR Morb Mortal Wkly Rep 1993;42(23):443–8.

Journal of HOSPITAL MEDICINE www.journalofhospitalmedicine.com

ORIGINAL RESEARCH

Post-Discharge Hospital Utilization Among Adult Medical Inpatients With Depressive Symptoms

Suzanne E. Mitchell, MD, MSc[1]
Michael K. Paasche-Orlow, MD, MA, MPH[2]
Shaula R. Forsythe, MPH[1]
Veerappa K. Chetty, PhD[1]
Julie K. O'Donnell, BA[1]
Jeffrey L. Greenwald, MD[3]
Larry Culpepper, MD, MPH[1]
Brian W. Jack, MD[1]

[1] Department of Family Medicine, Boston University School of Medicine/ Boston Medical Center, Boston, Massachusetts.

[2] Section of General Internal Medicine, Boston University School of Medicine/ Boston Medical Center, Boston, Massachusetts.

[3] Hospital Medicine Unit, Department of Medicine, Boston University School of Medicine/ Boston Medical Center, Boston, Massachusetts.

Dr. Mitchell's participation on this project was supported by a National Research Service Award (NRSA) Fellowship Grant. Support for Project RED is funded by the Agency for Healthcare Research and Quality.

The authors report no conflicts of interest related to this publication.

Dr. Suzanne Mitchell is a speaker on Relationship-centered Care and Cross-cultural communications for Manhattan Cross Cultural Group, Inc. and Merck & Co.

BACKGROUND: Little evidence exists to determine whether depression predicts hospital utilization following discharge among adult inpatients on a general medical service.

OBJECTIVE: We aimed to determine whether a positive depression screen during hospitalization is significantly associated with an increased rate of returning for hospital services.

DESIGN: A secondary analysis was performed using data from 738 English-speaking, hospitalized adults from the Project RED randomized controlled trial (clinicaltrials.gov Identifier: NCT00252057) conducted at an urban academic safety-net hospital.

MEASUREMENTS: We used the nine-item Patient Health Questionnaire (PHQ-9) depression screening tool to identify patients with depressive symptoms. The primary endpoint was hospital utilization, defined as the number of emergency department (ED) visits plus readmissions within 30 days of discharge. Poisson regression was used to control for confounding variables.

RESULTS: Of the 738 subjects included in the analysis, 238 (32%) screened positive for depressive symptoms. The unadjusted hospital utilization within 30 days of discharge was 56 utilizations per 100 depressed patients compared with 30 utilizations per 100 non-depressed patients, incident rate ratio (IRR) (confidence interval [CI]), 1.90 (1.51–2.40). After controlling for potential confounders, a higher rate of post-discharge hospital utilization was observed in patients with depressive symptoms compared to patients without depressive symptoms (IRR [CI], 1.73 [1.27–2.36]).

CONCLUSIONS: A positive screen for depressive symptoms during an inpatient hospital stay is associated with an increased rate of readmission within 30 days of discharge in an urban, academic, safety-net hospital population. ***Journal of Hospital Medicine* 2010;000:000–000.** © *2010 Society of Hospital Medicine.*

KEYWORDS: depression, hospital discharge, patient safety, readmission, rehospitalization.

Fully 19% of Medicare patients are readmitted to the hospital within 30 days of discharge.[1] This represents a large amount of potentially avoidable morbidity and cost. Indeed, projects to improve the discharge process and post-hospital care have shown that as much as one-third of hospital utilization in the month after discharge can be avoided.[2] Consequently, the rate of early, unplanned hospital utilization after discharge has emerged as an important indicator of hospital quality and the Centers for Medicare and Medicaid Services (CMS) has proposed a policy to decrease payments to hospitals with high rates of early unplanned hospital utilization. Thus, there is great interest in identifying modifiable risk factors for rehospitalization that could be used to refine intervention models and lead to improvements in quality of care, patient outcomes, and cost savings.

To date, known predictors of readmission include: lower socioeconomic status,[3] history of prior hospitalization[4] and advanced age,[5] length of stay greater than 7 days,[6] a high burden of comorbid illnesses (based on Charlson score),[7] poor social support,[8] and specific diagnoses (eg, congestive heart failure, chronic obstructive pulmonary disease [COPD] and myocardial infarction).[5,9,10] In addition, unplanned readmissions and emergency department (ED) visits have been linked to polypharmacy and adverse drug events related to treatment with medications such as warfarin, digoxin and narcotics.[11,12] Another characteristic that has also been linked to readmission is depression;[13] however[5]

2010 Society of Hospital Medicine DOI 10.1002/jhm.673
Published online in wiley InterScience (www.interscience.wiley.com).

reports supporting this association are from studies of elderly patients or with patients who have specific diagnoses (eg, congestive heart failure [CHF], COPD, myocardial infarction).[14–16]

Depression is common, affecting 13% to 16% of people in the US, and is recognized as an important risk factor for poor outcomes among patients with various chronic illnesses.[17–19] The mechanisms by which depression can be linked to health outcomes and health service utilization have been studied in age-specific or disease-specific cohorts such as cardiac patients or frail elders and include both physiologic factors such as hypercoagulability and hyperinflammatory conditions, as well as behavioral factors such as poor self-care behaviors and heightened sensitivity to somatic symptoms. How these mechanisms link depression to health outcomes and hospital utilization in a general medical population is not clearly understood. Kartha et al.[13] reported findings indicating that depression is a risk factor for rehospitalization in general medical inpatients, but the study sample was relatively small and the study design methodology significantly limited its generalizability.[12] It would be useful to provide supporting evidence showing depression as an important risk factor for readmission in the general medical in-patient population using more rigorous study methods and a larger cohort.

We hypothesized that depressive symptoms would be an independent risk factor for early unplanned hospital utilization after discharge for all medical patients. Therefore, we conducted a secondary analysis of the Project RED clinical trial dataset to assess the association between a positive depression screen during inpatient hospitalization and the rate of subsequent hospital utilization.

Methods

Data from the Project RED clinical trial were reviewed for inclusion in a secondary analysis. Complete data were available for 738 of the 749 subjects recruited for Project RED.

Project RED Setting and Participants

Project RED was a two-armed randomized controlled trial of English-speaking adult patients, 18 years or older, admitted to the teaching service of Boston Medical Center, a large urban safety-net hospital with an ethnically diverse patient population. A total of 749 subjects were enrolled and randomized between January 3, 2006 and October 18, 2007. Patients were required to have a telephone, be able to comprehend study details and the consent process in English, and have plans to be discharged to a US community. Patients were not enrolled if they were admitted from a skilled nursing facility or other hospital, transferred to a different hospital service prior to enrollment, admitted for a planned hospitalization, on hospital precautions, on suicide watch, deaf or blind. The Institutional Review Board of Boston University approved all study activities. A full description of the methods for the Project RED trial has been described previously.[2]

Outcome Variable

The primary endpoint was rate of hospital utilization within 30 days of discharge from the index admission, defined as the total number of ED visits and readmissions per subject within 30 days of the index discharge. Hospital utilization rates within 60 and 90 days of the index hospitalization discharge were also analyzed as secondary outcomes. Any ED visit in which a subject was subsequently admitted to the hospital was only counted as a readmission. Outcome data were collected by reviewing the hospital's electronic medical records (EMRs) and by contacting subjects by telephone 30 days after discharge. Dates of hospital utilization occurring at Boston Medical Center were obtained from the EMR, while those at other hospitals were collected through subject report. Subjects who could not be reached within 60 days of discharge were assumed alive.

Primary Independent Variable

The primary independent variable of interest was depressive symptoms defined as a positive score for minor or major depression on the nine-item Patient Health Questionnaire (PHQ-9) depression screening tool.[20] A dichotomized variable was created using a standardized scoring system to determine the screening cut-off for major or minor depressive symptoms.[19]

Statistical Analysis

Demographic and other characteristics of the subjects were compared by depression status (Table 1). Potential confounders were identified a priori from the available literature on factors associated with rehospitalization. These included age, gender, marital status, health literacy score (rapid estimate of health literacy in adult medicine tool [REALM]),[21] Charlson score,[22] insurance type, employment status, income level, homelessness status within past three months, hospital utilization within the 6 months prior to the index hospitalization, educational attainment, length of hospital stay and Project RED study group assignment. Bivariate analyses were conducted to determine which covariates were significant confounders of the relationship between depression and hospital utilization within 30 days of discharge. Chi-square tests were used for categorical variables and *t*-tests for continuous variables.

Age, length of stay, and Charlson score were used as continuous variables. Gender, marital status, frequent prior utilization (0–1 vs. 2 or more), and homelessness were treated as dichotomous variables. Categorical variables were created for, educational attainment (less than eighth grade, some high school, high school graduate, some college, college graduate), insurance type (Medicare, Medicaid, private insurance or free care), income level (no income, less than $10,000 per year, $10,000–20,000, $20,000–50,000, $50,000–

2010 Society of Hospital Medicine DOI 10.1002/jhm.673
Published online in wiley InterScience (www.interscience.wiley.com).

TABLE 1. Baseline Characteristics of Study Subjects by Depression Screen Status

Characteristic	Depression Screen* Negative (n = 500)	Positive (n = 238)	*P* Value
Race, No. (%)			
White	140 (30)	66 (30)	
Black	268 (58)	117 (54)	
Hispanic	47 (10)	29 (13)	0.760
Insurance, No. (%)			
Private	95 (19)	22 (9)	
Medicare	69 (14)	30 (13)	
Medicaid	214 (43)	143 (61)	
Free care†	118 (24)	40 (17)	<0.001
Education, No. (%)			
<8th grade	33 (7)	21 (9)	
Some high school	82 (17)	52 (22)	
High school grad	192 (38)	90 (38)	
Some college	126 (25)	51 (22)	
College grad	67 (13)	22 (9)	0.135
Health Literacy‡			
Grade 3 and below	64 (13)	44 (19)	
Grade 4–6	54 (11)	22 (10)	
Grade 7–8	156 (32)	73 (32)	
Grade 9 and above	213 (44)	89 (39)	0.170
Income, $, No. (%)			
No income	61 (12)	37 (16)	
<10K	77 (15)	61 (26)	
10–20K	96 (19)	35 (15)	
20–50K	97 (19)	34 (14)	
50–100K	35 (8)	7 (2)	
No answer	132 (27)	64 (27)	0.002
Employment status, No. (%)			
Full time	142 (28)	34 (14)	
Part time	57 (11)	30 (13)	
Not Working	297 (59)	171 (72)	<0.001
Age, mean (SD), years	49.9 (16.0)	49.6 (13.3)	0.802
Gender: No. (%) Female	239 (48)	133 (56)	0.040
Have PCP,§ No. (%) Yes	399 (80)	197 (83)	0.340
Marital status,‖ No. (%) unmarried	365 (73)	201 (85)	<0.001
Charlson score,¶ mean (SD)	1.058 (1.6)	1.56 (2.39)	0.001
RED study group,# No. (%)			
Intervention	243 (49)	127 (53)	0.22
Length of stay, days, mean (SD)	2.5 (2.8)	3.1 (3.8)	0.016
Homeless in last 3 months, No. (%)	45 (9)	30 (13)	0.130
Frequent utilizer,** No. (%)	159 (32)	104 (44)	0.002

NOTE: Some columns may not add up to 100% due to omission of "Other" categories.

Abbreviations: PCP, primary care provider; PHQ9, Patient Health Questionnaire-9; REALM, Rapid Estimate of Health Literacy in Adult Medicine tool; SD, standard deviation.

*Positive depressive symptom screen determined by PHQ9 screen tool, a nine-item 4-point Likert scale, standard scoring algorithm to screen for major and minor depression. A score of 5 or higher indicates a positive depression symptom screen.[17]

†Free Care is a Massachusetts state program for uninsured patients.

‡Health literacy categories correspond to total score as determined by REALM.[18]

§"Have PCP" refers to subject self-identifying PCP at time of Project RED study enrollment.

‖"Unmarried" marital status includes subjects identified as divorced, widow, single, partnered.

¶Charlson Comorbity Index Score reflects the cumulative increased likelihood of 1-year mortality. The higher the score the more severe the comorbid condition. A 33% increase in risk for death is reflected in a 1-point increase in weights. The minimum score is 0. There is no maximum score.[19]

#Project RED study intervention group refers to subjects who received the 3-armed discharge intervention.

**Frequent Utilizer is defined as a subject with 2 or more hospital utilizations in 6 months prior to Project RED clinical trial index admission.

100,000, no answer), level of health literacy (grade 3 and below, grade 4–6, grade 7–8, grade 9 or above) and employment status(working full-time, working part-time, not working, no answer).

The 30-day hospital utilization rate reflects the number of hospital utilization events within 30 days of discharge per subject. The same method was used to calculate hospital utilization rates within 60 and 90 days of discharge respectively. The unadjusted incident rate ratio (IRR) was calculated as the ratio of the rate of hospital utilizations among patients with depressive symptoms versus patients without depressive symptoms. Data for hospital utilization at 30, 60, and 90 days are cumulative.

Poisson models were used to test for significant differences between the predicted and observed number of hospitalization events at 30 days. A backward stepwise regression was conducted to identify and control for relevant confounders and construct the final, best-fit model for the association between depression and hospital reutilization. A statistical significance level of $P = 0.10$ was used for the stepwise regression. To evaluate potential interactions between depression and the Project RED intervention, interaction terms were included. Two-sided significance tests were used. *P* values of less than 0.05 were considered to indicate statistical significance. All data were analyzed with S-Plus 8.0 (Seattle, WA).

In addition, a Kaplan-Meier hazard curve was generated for the first hospital utilization event, ED visit or readmission, for the 30-day period following discharge and compared with a log-rank test.

Results

A total of 28% of subjects were categorized as having a positive depression screen. More women (36%) had positive depression screens than men (28%). Among patients with a positive depression screen, 58% had a history of depression and 53% were currently taking medications at the time of enrollment, compared with 25% and 22% respectively for subjects with a negative depression screen. Table 1 presents the means or percentages for baseline characteristics by depression status in the analytic cohort. Subjects with Medicaid for insurance had a higher rate of depression (61%) than subjects with Medicare (13%), private insurance (9%), or those who qualified for the Free Care pool (17%) which is the Massachusetts state funding for healthcare to uninsured persons. Subjects who were unemployed, unmarried, or who reported earnings less than $10,000 per year were also more likely to screen positive for depression. In addition, depressed subjects had a higher severity of co-morbid disease and longer length of stay for the index hospitalization. Patients categorized as frequent utilizers (2 or more prior admissions) for the 6 months prior to the index hospitalization were also more likely to be depressed. Of further note, is the relatively younger average age among both depressive

2010 Society of Hospital Medicine DOI 10.1002/jhm.673
Published online in wiley InterScience (www.interscience.wiley.com).

TABLE 2. Number of Hospital Utilizations, Hospital Utilization Rate, Unadjusted IRR at 30, 60 and 90 Days by Depression Screen Status*

	Depression Screen*			
Hospital Utilization	Negative, n = 500 (68%)	Positive, n = 238 (32%)	*P* Value	IRR (CI)
No. of hospital utilizations†	140	134		1.90 (1.51,2.40)
30-day hospital utilization rate	0.296	0.563	<0.001	
No. of hospital utilizations†	231	205		1.87 (1.55,2.26)
60-day hospital utilization rate	0.463	0.868	<0.001	
No. of hospital utilizations†	324	275		1.79 (1.53,2.10)
90-day hospital utilization rate	0.648	1.165	<0.001	

Abbreviations: CI, confidence interval; ED, emergency department; IRR, incident rate ratio.

*Depression screen determined by scoring of Patient Health Questionnaire-9 (PHQ9). Depressive symptom score of 5 points or higher is designated as positive.[17]

†Number of hospital utilizations include all ED visits and hospital readmissions following discharge from Project RED index admission. ED visits leading to hospital admission are counted as one event. Sum reflects cumulative number of events over 30, 60 and 90 days.

TABLE 3. Adjusted Incident Rate Ratio of Hospital Utilization Within 30 Days of Discharge

Characteristics	IRR	CI	*P* Value
Depression symptoms*			<0.001
Positive	1.73	1.27–2.36	
Negative	REF	1.0	
Gender			<0.001
Male	1.87	1.47–2.40	
Female	REF	1.0	
Marital status†			0.005
Married	0.625	0.44–0.89	
Unmarried	1.0	REF	
Frequent utilizer‡			<0.001
2+ prior visits	2.45	1.92–3.15	
<2 prior visits	1.0	REF	
Study group§			0.054
Intervention	0.76	0.55–1.06	
Control	1.0	REF	
Employment			
Part time	1.40	0.85–2.30	0.095
Not working	1.67	1.15–2.44	0.003
Other	0.52	0.07–3.85	0.262
Full time	1.0	REF	
Charlson Score‖	0.98	0.92–1.04	0.250
Group* depression¶	0.84	0.52–1.36	0.236
Age	1.00	0.99–1.01	0.375

Abbreviations: CI, confidence interval; ED, emergency department; IRR, incident rate ratio, PHQ9, Patient Health Questionnaire-9.

*Positive depressive symptom screen determined by PHQ9 screen tool, a nine-item 4-point Likert scale, standard scoring algorithm to screen for major and minor depression. A score of 5 or higher indicates a positive depression symptom screen.[17]

†Unmarried refers to subjects whose self-reported marital status includes divorced, single, partnered or widowed.

‡Frequent utilizer: 2 or more ED or hospital admissions visits in prior 6 months from index admission.

§Refers to Project RED study group assignment.

‖Charlson Comorbity Index Score reflects the cumulative increased likelihood of 1-year mortality. The higher the score the more severe the comorbid condition. A 33% increase in risk for death is reflected in a 1-point increase in weights. The minimum score is 0. There is no maximum score.[19]

¶Interaction term of Project RED study group assignment and depressive symptom category (positive or negative).

patients (49.6 years) and non-depressive patients (49.9) of these study subjects.

The unadjusted hospital utilization rate at 30, 60, and 90 days post-discharge by depression status is shown in Table 2. At 30 days post-discharge, those with depressive symptoms had a higher rate of hospital utilization than those without depressive symptoms (0.563 vs. 0.296). In other words, 56 utilization events occurred per 100 patients with depressive symptoms, compared with 30 utilization events per 100 patients without depressive symptoms. The unadjusted 30-day post-discharge hospital utilization rate among those with depressive symptoms was higher compared with those without symptoms (IRR, 1.90, 95% confidence interval [CI], 1.24–2.71). A similar trend was found among subjects at 60 and 90 days post-discharge.

Poisson regression analyses were conducted to control for potential confounding in the relationship between depressive symptoms and hospital utilization rate within 30 days after discharge (Table 3). After controlling for relevant confounders, including age, gender, employment status, frequent prior hospitalization status, marital status, Charlson score, Project RED study group assignment and the interaction variable for RED study group assignment and depression, the association between symptoms of depression, and hospital utilization rate remained significant (IRR, 1.73; 95% CI, 1.27–2.36).

Figure 1 depicts the Kaplan-Meier hazard curve generated for time to first hospital utilization, stratified by depression status. While 21% of participants without symptoms of depression had a hospital utilization within 30 days, fully 29% of participants with symptoms of depression had a hospital utilization within 30 days ($P = 0.011$).

Discussion

Our study shows hospitalized patients who screen positive for depressive symptoms are significantly more likely to have a hospital visit (emergency room or rehospitalization) within 30 days of discharge than those who do not screen positive for depressive symptoms among medical patients admitted to an urban, academic, safety-net hospital. These findings are consistent with, and extend, prior reports regarding depression and rehospitalization in specific populations (ie, geriatrics) and specific diagnoses (ie, cardiovascular disease [CVD] and COPD).[10–12] We observed a 73% higher incidence rate for hospital utilization within 30 days of discharge for those with symptoms of depression. This puts symptoms of depression on par with frequent prior rehospitalization, advanced age and low social support, as known risk factors for rehospitalization.[4,5,23]

2010 Society of Hospital Medicine DOI 10.1002/jhm.673
Published online in wiley InterScience (www.interscience.wiley.com).

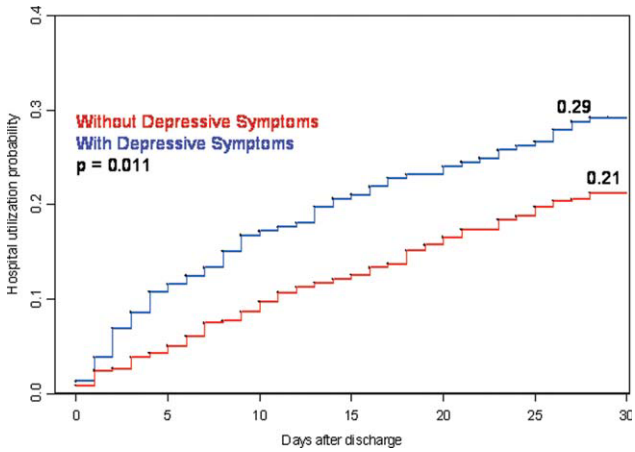


FIGURE 1. Hazard for hospital utilization among subjects with and without depressive symptoms in 30 days following hospital discharge.

Also of significance is the relatively young age of this study population (49.9 years non-depressive patients and 49.6 years for depressive patients) compared with the study cohorts used for research in the majority of the existing literature. The chief reason for the young age of our cohort is that potential subjects were excluded if they came from a skilled nursing facility or other hospital. This may limit the generalizability of our findings; however, it seems likely that interventions relating to depression and transitions of care will need to be quite different for patients that reside in long-term care facilities vs. patients that live in the community. For example, patients living in the community may have significant barriers to access post-discharge services due to insurance status and are more likely to be sensitive to variations in social support.

Early rehospitalization is associated with significant morbidity, mortality, and expense. It is also a potential marker for poor quality of care.[24] Concerns for patient safety, escalating healthcare costs, and possible change in hospital reimbursement mechanisms are fueling the search for modifiable risk factors associated with early rehospitalization. Our data provide evidence that symptoms of depression may be an important focus of attention. We do not know, however whether treating hospitalized patients who screen positive for depression will decrease early rehospitalization and emergency room utilization rates.

Various physiologic and behavioral mechanisms may link symptoms of depression to hospital utilization after discharge. For example, depressed patients with features of somatization may be more likely to experience worrisome physical symptoms after discharge and present prematurely for reevaluation. Patients who are sicker in some fashion not captured by our measured confounders may have symptoms of depression related to chronic, debilitating disease warranting early return to the hospital. Depression may also yield nonadherence to aspects of the discharge treatment plan leading to rehospitalization as a result of poor post-discharge disease management. For example, research shows that patients with depression following coronary artery bypass surgery are less likely to adhere with cardiac rehabilitation programs.[25] Likewise, depression among chronically ill patients such as diabetics, asthmatics, or human immunodeficiency virus (HIV)-positive patients impairs medication adherence and self-care behavior which may lead to disease relapse or recurrence.[26–28] One study examining depression effects on hypertensive medicine adherence in African Americans identified self-efficacy as a mediating factor between depression and nonadherence.[29] This implies that interventions such as self-management education, a program through which chronically-ill patients learn to better manage their illnesses through enhanced self-confidence and problem-solving strategies (including mood disorder challenges) may reduce early rehospitalization among depressed patients.[30]

There is also evidence that depression may have direct physiologic consequences. In patients with CVD, depression is associated with poor outcomes possibly related to decreased heart rate variability, hypercoagulability, high burdens of inflammatory markers, and severity of left ventricular dysfunction.[31–34] Similarly, depression among HIV/acquired immune deficiency syndrome (AIDS), diabetics and multiple sclerosis (MS) patients is linked to heightened levels of proinflammatory markers and less favorable outcomes that may signal a more severe form of the disease or an impaired response to treatment.[35–38] Indeed, MS investigators now hypothesize that the proinflammatory environment associated with the neurologic manifestations of MS are also causing depression symptoms among MS patients.[34] This theory contrasts the common belief that depression in the chronically ill manifests independent of the chronic illness or in response to living with chronic disease.

A major strength of the current study is the large dataset and the broad range of covariates available for analyses. However, several limitations should be noted. First, data on hospital utilization outside Boston Medical Center were determined by patient self-report and were not confirmed by document review. Second, we do not know the direction of the associations we report. If symptoms of depression are merely the consequence of having a higher disease burden, treatment of the underlying disease may be the most important response. While this is possible, our model does include several variables (eg, Charlson score and length of stay) that are likely to adjust for disease severity, pointing to the likelihood that symptoms of depression truly predict hospital utilization in a fashion that is independent of disease severity. Third, our results may not be generalizable to populations other than those served by urban safety-net hospitals or other populations excluded from the Project RED trial (eg, non-English speaking patients and patients from nursing homes). Finally, social factors such as substance use and social support system variables may residually confound the relationship between depression and hospital reutilization

2010 Society of Hospital Medicine DOI 10.1002/jhm.673
Published online in wiley InterScience (www.interscience.wiley.com).

demonstrated in this study. While this dataset does not include a measure of social support other than marital status and housing status, data is available on substance use. Analyses conducted by our colleagues using Project RED data found that in this study population depression was significantly more prevalent among substance users (29% vs. 14%) compared with non-users and that substance use is an independent risk factor for hospital reutilization (unpublished data).

Our findings linking depression to increased hospital utilization also warrant further consideration from healthcare policymakers. Central to the Obama Administration's February 2009 healthcare reform proposal is the pursuit of cost savings through reductions in unplanned hospital readmissions.[39] Thus, identifying potentially modifiable risk factors for readmission, such as depression, is of great concern to healthcare providers and policymakers across the nation. If, through testing of interventions, depression proves to be a modifiable risk for readmission, policymakers, while negotiating healthcare reform measures, must provide for the services required to address this comorbidity at the time of discharge. For example, if a patient screens positive for depressive symptoms during a hospitalization for COPD exacerbation, will the proposed payment reforms allow for mental health services during the immediate post-discharge period in order to reduce the likelihood of hospital readmission? Will those mental health services be readily available? Payment reforms that account for all necessary transitional care services will indeed help reduce readmission costs with less risk for untoward consequences.

In conclusion, our results indicate that a positive depression screen is a significant risk factor for early post-discharge hospital utilization among hospitalized adults on a general medical service, even after controlling for relevant confounders. Screening for depression during acute hospitalizations may be an important step in identifying patients at increased risk for readmission. Future research should focus on further characterizing and stratifying populations at highest risk for depression. Efforts should also include developing and evaluating targeted interventions for patients with symptoms of depression among hospitalized patients as part of discharge planning. Timely depression therapy during the hospitalization or following hospital discharge might reduce costly readmissions and enhance patient safety.

Address for correspondence and reprint requests:
Suzanne E. Mitchell, MD, MSc, Department of Family Medicine, Dowling 5, Boston Medical Center, 1 BMC Place, Boston, MA 02118; Telephone: 617-414-4465; Fax: 617-414-3345; E-mail: suzanne.mitchell@bmc.org Received 17 June 2009; revision received 8 December 2009; accepted 24 January 2010.

References

1. Jenks SF, Williams MV, Coleman, EA. Rehospitalizations among patients in the Medicare fee-for-service program. *N Engl J Med.* 2009;360(14): 1457–1459.
2. Jack BW, Chetty VK, Anthony D, et al. The reengineered hospital discharge program to decrease rehospitalization. *Ann Intern Med.* 2009; 150(3):178–187.
3. Weissman JS, Stern RS, Epstein AM. The impact of patient socioeconomic status and other social factors on readmission: a prospective study in four Massachusetts hospitals. *Inquiry.* 1994;31(2):163–172.
4. van Walraven C, Mamdani M, Fang J, Austin PC. Continuity of care and patient outcomes after hospital discharge. *J Gen Intern Med.* 2004;19: 624–631. [PMID: 15209600]
5. Marcantonio ER, McKean S, Goldfinger M, Kleenfield S, Yurkofsky M, Brennan TA. Factors associated with unplanned hospital readmission among patients 65 years of age and older in a Medicare managed care plan. *Am J Med.* 1999;107(1):13–17.
6. Krumholz HM, Parent EM, Tu N, et al. Readmission after hospitalization for congestive heart failure among Medicare beneficiaries. *Arch Intern Med.* 1997;157(1):99–104.
7. Librero J, Peiro S, Ordinana R. Chronic comorbidity and outcomes of hospital care: length of stay, mortality and readmission at 30 and 365 days. *J Clin Epidemiol.* 1999;52(3):171–179.
8. Rodríguez-Artalejo F, Guallar-Castillón P, Herrera MC, et al. Social network as a predictor of hospital readmission and mortality among older patients with heart failure. *J Card Fail.* 2006;12:621–627.
9. Wong AW, Gan WQ, Burns J, Sin DD, van Eeden SF. Acute exacerbation of chronic obstructive pulmonary disease: influence of social factors in determining length of stay and readmission rates. *Can Respir J.* 2008; 15(7):361–364.
10. Parashar S, Rumsfeld JS, Spertus JA, et al. Time course of depression and outcome of myocardial infarction. *Arch Intern Med.* 2006;166(18):2035–2043.
11. Budpitz DS, Shebab N, Kegler SR, et al. Medication use leading to emergency department visits for adverse drug events in older adults. *Ann Intern Med.* 2007;147(11):755–765.
12. Campbell SE, Seymour DG, Primrose WR. A systematic literature review of factors affecting outcomes in older medical patients admitted to hospital. *Age Ageing.* 2004;33(2):110–115.
13. Kartha A, Anthony D, Manasseh CS, et al. Depression is a risk factor for rehospitalization in medical inpatients. *Prim Care Companion J Clin Psychiatry.* 2007;9(4):256–262.
14. Almagro P, Barreiro Bienvenido, Ochoa de Echaguen A, et al. Risk factors for hospital readmission in patients with chronic obstructive pulmonary disease. *Respiration.* 2006;73:311–317.
15. Frasure-Smith N, Lesperance F, Gravel G, et al. Depression and healthcare costs during the first year following myocardial infarction. *J Psychosom Res.* 2000;48(4–5):471–478.
16. Jiang W, Alexander J, Christopher E, et al. Relationship of depression to increased risk of mortality and rehospitalization. *Arch Intern Med.* 2001; 161(15):1849–1856.
17. Parashar S, Rumsfeld JS, Spertus JA, et al. Time course of depression and outcome of myocardial infarction. *Arch Intern Med.* 2006;166:2035–2043.
18. Scherer M, Herrmann-Lingen C. Single item on positive affect is associated with 1-year survival in consecutive medical inpatients. *J Gen Hosp Psych.* 2009;31:8–13.
19. Hasin DS, Goodwin RD, Stinson FS, Grant BF. Epidemiology of major depressive disorder: results from the National Epidemiologic Survey on Alcoholism and Related Conditions. *Arch Gen Psychiatry.* 2005;62(10): 1097–106.
20. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: Validity of a brief depression severity measure. *J Gen Intern Med.* 2001;16:606–613. [PMID: 11556941]
21. Davis TC, Long SW, Jackson RH, et al. Rapid estimate of adult literacy in medicine: a shortened screening instrument. *Fam Med.* 1993;25:391–395. [PMID:8349060]
22. 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. [PMID: 3558716]
23. Rodriguez-Artalejo F, Guallar-Castillon P, Herrera MC, et al. Social network as a predictor of hospital readmission and mortality among older patients with heart failure. *J Card Fail.* 2006;12(8):621–627.

2010 Society of Hospital Medicine DOI 10.1002/jhm.673
Published online in wiley InterScience (www.interscience.wiley.com).

24. Ashton CM, Del Junco DJ, Souchek J, Wray NP, Mansyr CL. The association between the quality of inpatient care and early readmission: a meta-analysis of the evidence. *Med Care.* 1997;35(10):1044–1059.
25. Kronish IM, Rieckmann N, Halm FA, et al. Persistent depression affects adherence to secondary prevention behaviors after acute coronary syndromes. *J Gen Intern Med.* 2006;21(11):1178–1183.
26. Cukor D, Rosenthal DS, Jindal RM, Brown CD, Kimmel PL. Depression is an important contributor to low medication adherence in hemodialyzed patients and transplant recipients. *Kidney Int.* 2009;75(11):1223–1229.
27. Gonzalez JS, Safren SA, Delahanty LM, et al. Symptoms of depression prospectively predict poorer self-care in patients with Type 2 diabetes. *Diabet Med.* 2008;25(9):1102–1107.
28. Lima VD, Geller J, Bangsberg DR, et al. The effect of adherence on the association between depressive symptoms and mortality among HIV-infected individuals first initiating HAART. *AIDS.* 2007;21(9):1175–1183.
29. Schoenthaler A, Ogedegbe G, Allegrante JP. Self-efficacy mediates the relationship between depressive symptoms and medication adherence. *Health Educ Behav.* 2009;36(1):127–137.
30. Bodenheimer T, Lorig K, Holman H, Grumbach K. Patient self-management of chronic disease in primary care. *JAMA.* 2002;288(19): 2469–2475.
31. McFarlane AM. Effects of sertraline on the recovery rate of cardiac autonomic function in depressed patients after acute myocardial infarction. *Am Heart J.* 2001;142:617–623.
32. van Melle JP, de Jonge P, Ormel J, et al. Relationship between left ventricular dysfunction and depression following myocardial infarction: data from the MIND-IT. *Eur Heart J.* 2005;26:2650–2656.
33. Serebruany VL, Glassman AH, Malinin AI, et al. Platelet/endothelial biomarkers in depressed patients treated with the selective serotonin reuptake inhibitor sertraline after acugte coronary events: the Sertraline AntiDepressant Heart Attack Randomized Trial (SADHART) Platelet SubStudy. *Circulation.* 2003;108:939–944.
34. Mulvihill NT, Foley JB. Inflammation in acute coronary syndromes. *Cleve Clin J Med.* 2002;69(Suppl2):SII130–SII142.
35. Gold SM, Irwin MR. Depression and immunity: inflammation and depressive symptoms in multiple sclerosis. *Neurol Clin.* 2006;24(3):507–519.
36. Brydon L, Walker C, Wawrzyniak A, et al. Synergistic effects of psychological and immune stressors on inflammatory cytokines and sickness responses in humans. *Brain Behav Immun.* 2009;23(2):217–224.
37. Gresson JM, Hurwitz BE, Llabre MM, Schneiderman N, Penedo FJ, Klimas NG. Psychological distress, killer lymphocytes and disease severity in HIV/AIDS. *Brain Behav Immun.* 2008;22(6):901–911.
38. Pizzi C, Manzoli L, Mancini S, Costa GM. Analysis of potential predictors of depression among coronary heart disease risk factors including heart rate variability, markers of inflammation, and endothelial function. *Eur Heart J.* 2008;29(9):1110–1117.
39. Connolly C. Obama proposes $634 billion fund for health care. *Washington Post.* February 26, 2009:A1.

2010 Society of Hospital Medicine DOI 10.1002/jhm.673
Published online in wiley InterScience (www.interscience.wiley.com).

Treating and Precepting with RESPECT: A Relational Model Addressing Race, Ethnicity, and Culture in Medical Training

Carol Mostow, LICSW[1], Julie Crosson, MD[2,9], Sandra Gordon, MD[3,9], Sheila Chapman, MD[3,9], Peter Gonzalez, MD[10,11], Eric Hardt, MD[6,9], Leyda Delgado, MD[3,8,9], Thea James, MD[7,9], and Michele David, MD, MPH, MBA[3,4,5,9]

[1]Department of Family Medicine, Boston Medical Center, Boston, MA, USA; [2]Dorchester House Multi-Service Center, Dorchester, MA, USA; [3]Department of Medicine, Section of General Internal Medicine, Boston Medical Center, Boston, MA, USA; [4]Boston University Center for Excellence in Women's Health, Boston, MA, USA; [5]Haitian Health Institute at Boston Medical Center, Boston, MA, USA; [6]Department of Medicine, Section of Geriatrics, Boston Medical Center, Boston, MA, USA; [7]Department of Emergency Medicine, Boston Medical Center, Boston, MA, USA; [8]Latino Clinic, Boston Medical Center, Boston, MA, USA; [9]Boston University School of Medicine, Boston, MA, USA; [10]Harvard Medical School, Boston, MA, USA; [11]Beth Israel Deaconess Medical Center, Boston, MA, USA.

BACKGROUND: In 2000 a diverse group of clinicians/educators at an inner-city safety-net hospital identified relational skills to reduce disparities at the point of care.

DESCRIPTION: The resulting interviewing and precepting model helps build trust with patients as well as with learners. RESPECT adds attention to the relational dimension, addressing documented disparities in respect, empathy, power-sharing, and trust while incorporating prior cross-cultural models. Specific behavioral descriptions for each component make RESPECT a concrete, practical, integrated model for teaching patient care.

CONCLUSIONS: Precepting with RESPECT fosters a safe climate for residents to partner with faculty, address challenges with patients at risk, and improve outcomes.

KEY WORDS: health care disparities; cultural competency training; cross-cultural medicine; learning climate; physician-patient relationship; communication; professionalism.
J Gen Intern Med
DOI: 10.1007/s11606-010-1274-4

INTRODUCTION

The need to reduce health care disparities at the point of care raises urgent challenges for training of medical professionals. Each clinical encounter offers an opportunity to decrease risk for disparate care and poor outcomes. To succeed in this effort, clinicians must transform knowledge and awareness into action and skills; we must partner effectively with patients whose backgrounds may differ dramatically from our own. An expanding literature has identified patient-centered and culturally competent interpersonal skills as critical to maximizing patient satisfaction, trust, adherence, and outcomes.[1] The challenges for clinical faculty are twofold. How do we establish the trust and safe learning climate trainees need to discuss openly their challenges in interactions with patients at risk for disparities? Secondly, if our learners do identify barriers to effective relationships with their patients, what skill set should we teach to help them be more effective? The well-documented decline in medical trainees' empathy during the clinical years means that house staff may be at particular risk for erosion of the interest and personal skills essential to maximize their patients' care[2,3]. Attention to learning climate and supportive training relationships may help restore some of residents' empathy, while skills training remains important given the often pragmatic orientation of trainees and the need to sustain their motivation and self-efficacy.

In 2000 a racially and culturally diverse group of clinician-educators at Boston Medical Center, a large urban safety net hospital, worked to address this training need, identifying the communication skills we use to relate effectively with our patients, a population with great diversity by race/ethnicity, culture, language, and class. We sought ways to address disparities within each individual doctor-patient encounter and new methods to teach these skills to busy trainees in clinical settings. The resulting RESPECT model is an integrated instrument for teaching, evaluation, and expansion of faculty skills. It provides a guide for what to do and how to teach it. By addressing relations with learners as well as patients, it offers help aligning the powerful hidden curriculum with the explicit curricular goals of compassionate patient care, sensitivity to diverse populations, interpersonal communication, professionalism, life-long learning, and

This article is dedicated with gratitude and respect in memory of Samuel Morse Putnam MD, MPH.

 The RESPECT model was developed by members of the Diversity Curriculum Task Force in 2000 originally led by Carol Mostow, Peter Gonzalez, and Julie Crosson including Sandra Gordon, Sheila Chapman, Oranti Aladessamni, Thea James, Eric Hardt, Charles Brackett, and Sandhya Wahi. The model was further elaborated and implemented in the 2004-2005 faculty development project, which in addition to Mostow, Gordon and Crosson included Leyda Delgado, Michele David, Sondra Crosby, Danru Lee, and Sam Putnam.

practice-based improvement. These are all core residency training requirements for accreditation by the American Council of Graduate Medical Education (ACGME)[4].

What is the RESPECT Model?

The RESPECT model is an action-oriented set of communication and relational behaviors designed to build trust across differences of race/ethnicity, culture, and power[5]. Clinicians can use RESPECT to target documented disparities in doctor-patient relationships[6]. Medical educators can use the model as an integrated instrument for teaching, observation, and evaluation. Finally, preceptors can also apply the RESPECT model to their own relationships with learners to facilitate more effective training. The specific component skills for this interviewing model and educational framework are:

*R*espect
*E*xplanatory model
*S*ocial context, including Stressors, Supports, Strengths and Spirituality
*P*ower
*E*mpathy
*C*oncerns
*T*rust/ *T*herapeutic alliance/ *T*eam.

See Table 1 for details on the use of RESPECT for patient care.

Why Another Model?

A review of the literature reveals important early contributions regarding how to elicit and negotiate differences of culture, health beliefs, and practices. Valuable attention to social context helps to understand some issues of the poor, but we did not find the same attention to matters regarding race, power, and distrust apparent to our racially/ethnically diverse patients and clinician group. We also noted that teaching models initially focused on the traditional information-gathering function of the interview rather than a relational guide addressing the affective domain. In the context of an emotional climate of potential distrust, racial, ethnic, and economic inequities, and risk for disparate outcomes, we were looking for a model that teaches physicians how to respond, not just what to ask. What should the physician do to build trust across differences of power, race, ethnicity, and class? Finally, we knew as preceptors that these issues felt sensitive and difficult to address in busy clinical settings with harassed residents trying to handle patients with challenging medical as well as social problems. What in the literature could help us build the relationships with residents that allowed these important conversations, so important to help them better meet the needs of the patients they serve?

A review of the literature on cross-cultural interviewing identified Kleinman's tool to elicit diverse health beliefs and practices. This widely used set of questions targets the patient's explanatory model, concerns about the illness, and goals of treatment[7]. Building on this work, Berlin and Fowkes proposed the LEARN Model, a framework for cross-cultural care[8]. This model contributes the skills of *L*istening, *E*xplanation of provider perceptions, *A*cknowledgement of differences, treatment *R*ecommendations, and *N*egotiation of plans. The *ETHNIC* mnemonic, designed by Levin et al., adds helpful attention to the way that cultural beliefs and practices may differ and be addressed. Model components include: seeking *E*xplanations of the patients' understanding of illness; asking what *T*reatments patients use and expect; inquiring about alternative *H*ealers; *N*egotiation of options incorporating patients' beliefs; designing culturally appropriate *I*nterventions; and *C*ollaboration with patients and their support systems[9]. Betancourt et al. developed the *ESFT* model, which added to the *E*xplanatory model the expanded *S*ocial context of illness including economics, literacy, and other important constraints[10]. This model also included attention to the patient's *F*ears and concerns about therapy and advocates *T*herapeutic contracting with patients around medication issues.

These valuable models all facilitate the gathering of relevant information related to cultural issues and introduce the skill of negotiation. However, they focus more on negotiating difference than addressing the power dynamics involved with differences of race, ethnicity, and class, or the relational and affective dimensions of the clinical encounter. These models do not themselves suggest or describe direct provider responsiveness to the emotional climate and potential distrust often operative in encounters across differences of race, ethnicity, and power in our society.[11]

Since 2001 additional cross-cultural interviewing models have been developed, including the *BELIEF* model by Dobbie et al., that do recognize the importance of addressing the affective domain in conversations across cultures.[12] Its components are: health *B*eliefs, *E*xplanation, *L*earn, *I*mpact, *E*mpathy, and *F*eelings. Still, little attention is given to the dynamics of power and historic oppression important for addressing differences of race, ethnicity, and class in contemporary America. According to a recent critique of multicultural education, too many cultural competence curriculae fail to provide direction in these dimensions.[13]

In 2008, scholars conversant with the literature regarding cultural competence and patient-centered communication called for the two fields to build on each others' contributions to increase effectiveness in both domains[14]. Teal and Street furthered this effort by publishing a comprehensive model using a complex array of communication skills.[15] Their important work further supports the elements distilled and prioritized in the RESPECT model while suggesting curricular elements for an in-depth and intensive physician training process.

Why RESPECT the Patient?

Reduction of disparities in clinical practice requires strategic action to bridge differences of race/ethnicity, culture, and power. The RESPECT model can help medical professionals develop the practical skills needed to build trust actively. It expands on earlier models by adding the components of respect, empathy, and power specifically to target documented areas of disparity in interracial and cross-cultural encounters.[16] In the context of inequity, stigma, and power differentials, proactive demonstrations of respect seek to mitigate patients' prior negative experiences and possible expectations of disrespect.[17] In the RESPECT model, the provider actively conveys empathy to the patient rather than simply collecting data. Since so much of trust-building is affective and relation-

Table 1. How (and why) to RESPECT the Patient

Skill	Definition	Behavioral description	Examples	Relevant evidence
Respect: *show*	A demonstrable attitude communicating the value and autonomy of the patient and the validity of his/her concerns	Non-verbal: Maintain attentive posture, appropriate eye and personal contact; follow cues regarding personal space, physical contact, and appropriate greetings	"Hi, I'm Dr. X, and I'm looking forward to working with you"	Disparity: African American, Hispanic, and Asian patients reported feeling less respected by their doctors than did White patients[17]
		Verbal: Welcome patient to encounter; introduce self and explain role on team; ask the patient how they want to be addressed; recognize and affirm strengths and efforts	"What would you like me to call you?" "You overcame a lot to get here today!"	
Expanatory model: ask	The patient's understanding of what causes their illness, or what will help it	Non-verbal: Give patient space to share their ideas by listening without judgment Verbal: Ask patient what they think is causing or will alleviate their symptoms	"What do you or your family think is causing your symptoms?" "Why do you think this started when it did?" "What do you think will solve the problem?"	Patients and doctors often have different ideas that remain unexplored unless elicited. Without discussion, patients leave less satisfied[20]
Social context: ask	Impact of patient's life upon illness and of illness on his/her life. Include stressors, supports, strengths, spiritual resources that influence patient, health or care	Nonverbal: Show interest and pay attention Verbal: Ask how patient's illness affects their life and how their life affects illness	"What should I know about you to care for you best?" "What is hardest for you?" "Who helps you the most?" "What keeps you going?" "What about religion?"	Low social support predicts higher mortality post MI [21] Negative health consequences follow death of spouse alleviated by presence of confiding figure[22]
Power: *share*	Access to status, control, resources, options, and ability to produce desired outcomes Power gradient favors doctors[11]	Non-verbal: Reduce physical barriers. Don't dominate the interaction. Sit Verbal: Listen. Limit interruptions. Build history rather than take it.[24] Use EMR to share information with patient via graphs, etc. Invite open discussion of disagreement. Negotiate agenda/treatment plan by eliciting preferences. Empower patient, recognize strengths	"Beside your diabetes, what else should we talk about?" "What would make your medications easier?" "Thanks for telling me that you don"t agree. What do you think?"	Disparity: White physicians dominate conversation more with non-white patients[36] Self-efficacy needed to make healthy choices[23]
Empathy: *show*	Verbal and nonverbal responses that validate patients' emotions and cause them to feel understood.	Non-verbal: Listen attentively and respond accordingly Verbal: Name and validate patients' emotions. Put significance of patient's experience into words to convey specific understanding	"That must be hard, anyone would feel that way." "This can be scary. Let's talk about it." "The injury changed everything for you"	Disparity: Doctors display less warmth with African-American patients[36]
Concerns/ fears: ask	Worries about symptoms, diagnosis, or treatment, often unexpressed	Nonverbal: Head nods, etc., to encourage patient to give details Verbal: Ask open-ended questions about fears/concerns	"What worries you the most?" "What scares you about the medication?" "Are you worried about sex after your heart attack?"	Unvoiced concerns lead to unmet needs and patient dissatisfaction[25]
Trust/team-building/ therapeutic alliance	Relationship built on understanding, power-sharing and empathy; patient confident that doctor acts on his behalf	Trust: Notice/respond to signs of distrust. Elicit and respond to expectations. Reassure and clarify follow-up. Follow through	"People in my family have had the same thing." "Should we get your family involved to help us?" "We"re here when you need us." "Let's make sure we answer all your questions so you feel comfortable making your decision"	Disparity: 62.8% Blacks vs. 38.4% Whites believe their doctors have or would experiment on them without their consent.[26] Black patients receive less support, partnering and information[27]
Build: don't assume		Therapeutic alliance: Find specific common goals, negotiate differences Team building: Identify, enlist, and collaborate with potential members of health care team		

al, these empathic skills are particularly important.[18,19] Providers need skills to counterbalance power differentials through verbal and nonverbal behaviors, actively seeking patients' concerns, potential disagreements, and barriers to treatment.

Accumulating evidence supports the validity of the RESPECT components as critically related to disparities in care (see Table 1: refs [16,20–27,36]). For example, disparities were identified in patients' experience of respect and trust in surveys of patients and providers conducted by the Kaiser Family Foundation and the Commonwealth Fund.[17,28–34] Trust in doctors among African-American patients continues to be diminished by the legacy of the Tuskegee experiment, often reinforced by more contemporary racial experiences.[35,36] The components of power and empathy are also supported by evidence regarding disparate experiences of non-white patients. White physicians dominate speech more with non-white patients and display less warmth and patient-centered behaviors.[16,36] Another study documented that black patients received less support, partnering, and information from their doctors and have lower levels of trust in their physicians.[27]

Collaborative empowerment and empathy assume even greater importance given recent findings that physician dominance reduced patient engagement while demonstrations of caring elicited more emotional sharing by the patient and increased activation in the interview[37]. Another study found that the majority of patients do have the explanatory models of illness described by Kleinman as so pivotal in health-related behaviors, but did not disclose them unless specifically invited to share them.[20]

The contribution of social factors to health and patients' approach to health care has also been well established.[38] The RESPECT model offers a helpful tool to focus quickly and prioritize elements of the social context most salient to the patient. Physicians can focus on matters of greatest importance to the patient rather than squandering time on less relevant closed-ended demographic information. (see Table 1). The value of identifying the patient's stressors is supported by clinical studies.[39,40] By eliciting strengths, supports, and spiritual resources as well as stressors, the RESPECT model replaces a focus on social deficits with an appreciation of the internal, family, and community strengths fostering self-efficacy, a critical element for health behavior change.[23] Recent literature on patient-centered approaches suggests that prompt attention to patients' emotions and perspective may add to, not diminish efficiency.[41]

Why RESPECT the Trainee?

Recognizing the learning climate and working to improve teaching relationships addresses the implicit "hidden curriculum" within medical education.[42–44]. Much learning occurs by selective reinforcement of behaviors that may not be consciously examined.[45] Effective teaching must address the persisting decline in empathy documented during the clinical years of medical school and residency.[2,3,46] Caring for the socially disadvantaged adds to the personal stressors of residency training.[47] Explorations of residents' errors suggest a reluctance to discuss them with faculty.[48] The Institute of Medicine reports on patient safety indicate that rigid hierarchies and atmospheres of fault-finding and blame are antithetical to a safe culture for patient care. A "flattened hierarchy" promotes honesty and reduces defensiveness.[49] Growing attention to the implicit curriculum suggests the importance of using methods to train residents that are congruent with learning objectives and content.[50,51]

Our explicit application of the RESPECT model to both patients and learners is similar to an innovative contribution by Kern et al. teaching residents an approach to psychosocially sensitive care of patients while applying the same approach to the residents.[52] As they note, empathic listening skills may wither in residents whose own concerns are routinely ignored. A similar call to address power dynamics in the teaching of prospective doctors recently came from educators focused on the elimination of health disparities.[13] We must create a safe learning climate for the open dialogue that learners need to internalize professional values and commit to social justice goals. The call for a wholesale "reorientation of the traditional teacher-student paradigms" requires faculty development related to these new teaching methods.

RESPECT: A Versatile Teaching Tool

Ten years using RESPECT has persuaded us that it is an effective model promoting generalizable skills for a wide range of health care providers working in varied clinical settings. In addition to its utility in clinical practice, the RESPECT model has proven to be a versatile tool for teaching, evaluation, and faculty development.

For pre-clinical learners, didactic presentations and experiential approaches provide exposure to the relational process of clinical care. We adapted RESPECT as an observational tool for students to observe their attendings at work with patients (Looking for RESPECT, BOX 1). Following a lecture introducing the framework and an exercise about difference and power, all first year medical students at Boston University School of Medicine use the tool in their first patient contacts as they observe faculty in diverse clinical settings. They use Looking for RESPECT to guide their observations and write-ups of physician-patient encounters with a focus on patient concerns and relational elements beyond the classical data-oriented approach. Through the lens of the model, students see the clinical relationship and watch doctors build trust and understanding across differences.

Looking for RESPECT:

An observation exercise or assessment tool for demonstrating ACGME competencies

The RESPECT model conveys a value essential to clinical practice and required for all effective communication. It also offers a handy list of essential components to optimize health care encounters for both patient and physician, especially important when facing additional differences of culture, race, ethnicity, class, etc. between the parties.

R	Did the physician convey Respect? If so, in which ways? What seemed to be the patient's reaction?
E	What did the doctor learn about the patient's Explanatory model of the illness and ideas of what might help? (e.g. what's wrong with me? what do I think will make me better?)
S	What did the doctor learn about the Social context of the patient including Stressors, Supports, Strengths, Spiritual beliefs and practices which might affect health, treatment, expectations, or relationships with the doctor and healthcare system? How does the patient's life affect his illness? How does his illness affect his life?
P	What expectations and preferences did the patient and doctor each seem to have for the Power relationship? Notice both non-verbal and verbal cues. Pay attention to displays of deference, control, hierarchy. Who does the talking? Who determines the agenda ? What did the doctor or patient do or say during the encounter which seemed to empower or disempower either party? Does the doctor seek the patient's input and preferences or simply announce to the patient what will happen?
E	What opportunities were there for the doctor to convey Empathy? What did the doctor do or say to which conveyed an understanding of the patient's experience and its significance to the patient? What were the patient's responses?
C	Are there any Concerns and fears which underlie or coexist with the patient's presenting problems? What prompted the patient to share these? What did the physician do to elicit or facilitate this fuller, possibly more emotional disclosure?
T	What seemed to be the patient's initial level of Trust? What did the doctor say or do during the interview which seemed to modify the patient's level of Trust? What verbal and non-verbal indicators did you notice? Was a Therapeutic plan created? Did the doctor and patient reach common ground regarding the problem and the approach to diagnosis and treatment? Are there any indications that there might be obstacles to and/or disagreements about the next steps? If there remained divergent preferences at the end of the interview, was the doctor able to negotiate a partnership based on other shared goals?

NOTE: Make sure to observe non-verbal as well as verbal behaviors regarding the above.

To teach these concepts to residents and practicing clinicians, we use case-based learning in discussions, role plays, and trigger tapes. This approach has worked well in interactive didactic presentations with medical residents and clinicians in a variety of settings. As the model is easily generalized, faculty have used it to train substance abuse screeners and health care workers providing breast cancer outreach.[53] It has also been adopted for the didactics, discussion, and video demonstration for Alcohol Clinical Teaching, a web-based alcohol training program.[54] In a training for experienced eye care professionals and faculty, we solicited problematic cross-cultural clinical encounters cases and addressed them using RESPECT-based role play.[55]

For residents, we use an inductive exercise designed to help them discover the RESPECT model actively. In case-based preclinic conferences at continuity clinics, we encourage them to think through their own cases with challenging communication issues. They then identify the elements contributing to a past success bridging differences and distrust, which combine to form the RESPECT model. The residents then reflect and strategize with peers and faculty about how to apply the model to their other challenging cases, identify skills and behaviors to try in future encounters.

We have found the observation tool designed for medical students, Looking for RESPECT (BOX 1), to be equally useful for preceptors in resident primary care clinics. The model provides a framework that helps faculty go beyond evaluation of standard bedside skills to assess critical competencies of interpersonal communication and professionalism including sensitivity to the needs of diverse populations[4]. We can use the same teaching tool to stimulate residents' reflections and self-assessments as well as for more structured performance evaluation as has been suggested.[56]

Effective teaching in this area requires that residents disclose the challenges that they face in order to reflect on what they're doing and consider new approaches. A top-down approach is unlikely to succeed in promoting development of this level of 'critical consciousness.'[13] Simply telling learners to be culturally

sensitive and respectful does not assure that they can and will do so. We realized that we needed to apply the RESPECT model skills to build relationships with our learners as well as with our patients. We have translated those skills into a precepting guide that helps faculty improve their skills. By explicitly addressing issues of power, Precepting with RESPECT (BOX 2) addresses the implicit curriculum and harnesses parallels between educational and clinical encounters.[52] The model incorporates relational and cognitive elements from recommended precepting models that elicit the resident's model of understanding, encourage independent thinking, and facilitate targeted educational interventions.[57,58]

Precepting with RESPECT:

Application of the RESPECT model to preceptor-learner relationships

RESPECT	Approach the learner with respect. Builds resident confidence, preceptor resident relationship. Reduces defensiveness. "I know how hard you've been working to try and get his diabetes under control".
EXPLANATORY MODEL	Elicit the resident's thoughts about the patient and interest in the patient's perspective. Helps preceptor learn what resident knows and has asked as starting point for further discussion. Conveys interest in resident's perspective while supporting resident's interest in patient's perspective "What do you think is going on with the patient?" "What does patient think is causing his symptoms?"
SOCIAL CONTEXT	Check re residents' well-being and context . Explore possible professional and personal stressors. Builds preceptor-trainee relationship. Models how to act with patients. " How are things going for you these days?" "What clinical rotation are you on?"
POWER	Find ways to share power and support resident self-efficacy. Resist the temptation to take over in the face of learner's uncertainty. Helps assess clinical judgment, build problem-solving ability and increase investment in solution. "How might you find out why this patient has a hard time taking his medicines daily?"
EMPATHY	Let resident know their frustrations and emotions are heard Documented decline in residents' mood and empathy for patients as well as differences from their patients' background make this step essential. Faculty support may enable residents to engage more effectively with patient. "After all your effort, I can imagine how frustrating it is that she didn't fill her prescription." "Particularly when I'm tired, it can be hard to put aside my own frustration to find out what's going on from the patient's perspective."
CONCERNS	Elicit and address residents' concerns about situations they don't feel confident handling or fear will make visit too long. Help residents strategize about possible solutions and educate about relevant data Replaces anxiety with information to improve quality and efficiency " I know you were worried about eliciting more of the patient's concerns but eliciting them doesn't mean you or the patient can tackle them all in one visit." " Let's discuss how to identify your and the patient's top priorities, come up with a plan for today and bring him back for follow-up and to address the other issues."
TRUST	Building on all the skills above fosters trust in the preceptor-resident relationship. Learners may become more willing to identify areas of challenge. "I admire your openness and ability to share with me that your patient was so frustrated that she wanted to change doctors." "How can I help you?"

We discovered some of these approaches during a year-long faculty development project at our institution in 2004-2005, funded by BCBS of Massachusetts Medical Foundation.[59] We applied the RESPECT model in resident clinics and other settings. We sought to make cultural competence, prevention of health care communication disparities, and more effective relationships between residents and patients a routine part of precepting and clinical care. Faculty sought increased comfort and effectiveness precepting residents to connect more effectively with their diverse patient population. A group of ten experienced clinician-educators of diverse racial/ethnic backgrounds joined together in monthly 3-h sessions. The faculty group included four blacks, four whites, one Asian, and one Latina; four participants were born outside of the USA. Participants included clinic preceptors, in-patient attendings, an emergency room attending, and a medical educator, all involved with resident and/or medical student teaching. Most of them had participated in the ongoing Diversity Curriculum Task Force that had originally developed the RESPECT model. Despite our extensive experience, like our learners, we needed the opportunity to reflect and develop skills on how best to approach sensitive subjects and resident interactions that made us uncomfortable. While our commitment to the goals of cultural competence was clear, how to be most effective with our learners in getting there was not. Using Looking for RESPECT (BOX 1), preceptors identified residents' learning needs and then designed interventions for individual residents using Precepting with RESPECT (BOX 2). Preceptors often role-played these interactions to practice and get feedback from other faculty using faculty development strategies developed by the American Academy on Communication in Healthcare.[60] This protected learning environment allowed us to experiment with teaching methods, strategize about challenging teaching encounters, and share resulting observations of resident behaviors and attitudes.

Evaluation

In an effort to test, refine, and disseminate our RESPECT model into the clinical setting, we developed a self-administered questionnaire to evaluate its impact on teachers and trainees. In 2004-2005, we surveyed a diverse group of internal medicine residents at our hospital before and after implementing the RESPECT model as a teaching tool in resident primary care clinic sessions. Ten precepting faculty underwent structured interviews at the end of the intervention. Forty-four residents completed the baseline questionnaire and 20 the post-intervention questionnaire.

The ten precepting faculty all reported increased comfort and skill precepting regarding cross-cultural care in resident continuity clinics. In particular, some expressed a willingness to address racial issues affecting patient care that they had previously regarded as "taboo." Another theme noted was that faculty members also improved their own interviewing skills. Among the residents surveyed, 88% felt that the RESPECT-based training was useful. We found differences in several survey items pre- and post-training. Residents reported greater comfort when interacting with diverse groups (60% pre, 80% post). Residents seemed more convinced that cross-cultural training might improve health care delivery (57% pre, 93% post). Some reported major improvement in skills related to culturally sensitive interviewing (10% pre, 20% post). The number of residents reporting no skills in this area declined (12% pre, 0% post). Small numbers and limited response rates precluded more statistical analysis of survey results. Changes in preceptor and resident schedules along with variation in the frequency of contact also limited the data.

DISCUSSION

The RESPECT model was a product of a highly diverse group of clinician-educators working together for 10 years at a large urban medical center that cares for racially and ethnically diverse patients from the US and abroad. It is notable that diversity of the faculty group itself allowed us to draw upon personal and clinical experience and to share insights about the role of differences and of race in particular. Together we were able to derive a model of broad clinical utility to the care of our patients. A diverse faculty group's access to clinical and personal experience outside of majority culture may foster insights less apparent to others[28,62]. However, in the 2004–2005 evaluation of the project to improve faculty participants' teaching skills by precepting with RESPECT, our faculty preceptors also noted with interest improved interactions with their own patients by using the RESPECT model. The relational elements in the RESPECT model for interviewing and teaching have a lot to offer even a very sophisticated faculty and remind us that cultural competence and relationship-centered care of at-risk populations require ongoing efforts.

The RESPECT model can be useful to medical schools, residencies, and other health care organizations in meeting their training goals, core curricular competencies, and patient care standards, assuring that quality care is delivered equitably to all. To monitor and continually improve the quality of care and communication with all patients, evolving definitions by the ACGME[4] and others[44,63–65] of professionalism, practice-based improvement, and systems-based practice remind us that faculty and residents alike need to be lifelong learners. We must remain ever mindful of our ongoing impact on others and the value of teaming up with our patients, trainees, and colleagues. Respectful communication across power differentials among faculty, residents, and colleagues promotes positive professional communication and teamwork along with full and open disclosure.[49] This change of culture can promote a personal and systems-based examination of obstacles to optimum care and patient safety, especially important for patients at risk for poor outcomes.

One of the limitations of the RESPECT evaluation is that because of small resident numbers in the pilot clinics and the large number of them unavailable to complete the post-intervention survey, we could not perform tests of statistical difference. However, to date one study of the efficacy of the teaching model has been published.[55] Cultural competence training based on the RESPECT model was conducted for faculty at The New England Eye Institute and New England College of Optometry. A follow-up study demonstrated measurable positive impact that persisted 3 months post-intervention, using a cultural competency scale validated by Campinha-Bacote.[61]

It has been 10 years since we initially derived our interviewing and teaching model from the insights of our racially, ethnically diverse group of clinician educators. During that time empirical data as well as consensus recommendations

from the fields of patient-centered communication and cultural competence have been published that support the elements in our model.[1,6,15,16] For patients already at risk for disparities, a doctor who dominates the encounter or withholds warmth or empathy is a threat and a lost opportunity to partner for better health.[5,6,36] By addressing barriers to trust posed by power differentials, RESPECT may be a helpful part of the solution. What we offer is an easily remembered mnemonic that prioritizes elements linked to documented disparities in a clear, readily applied model. Attention to the power dynamics within our own relationships with residents has helped us precept our learners with RESPECT. Designed to improve communication with patients at risk for disparities, the RESPECT model provides a flexible, practical, and widely applicable teaching model targeting documented disparities in health care communication while promoting relational skills needed to build trust and partner for better health care.

Acknowledgements: *This work was supported in part by grants from the Kenneth B. Schwartz Center and by Blue Cross Blue Shield of Massachusetts Medical Foundation. We would also like to acknowledge the value to our work of Elaine Pinderhughes' attention to power and race (Understanding Race, Empathy and Power 1989), the focus on patient and relationship-centered care by the American Academy on Communication in Healthcare (formerly AAPP), and the LEARN and ESFT models cited below. Aileen Richmond, MPH, provided valuable assistance with data collection, faculty interviews, and collation of themes. We also thank Beth Lown, MD, for kindly reviewing the manuscript.*

Conflict of Interest: *None disclosed.*

Corresponding Author: *Carol Mostow, LICSW; Department of Family Medicine, Boston Medical Center, 771 Albany Street, Dowling 5, Boston, MA 02118, USA (e-mail: Carol.Mostow@bmc.org).*

REFERENCES

1. **Fuertes JN, Boylan LS, Fontanella JA.** Behavioral indices in medical care outcome: the working alliance, adherence and related factors. JGIM. 2009;24(1):80–5.
2. **Bellini LM, Baime M, Shea JA.** Variation of mood and empathy during internship. JAMA. 2002;287(23):3143–6.
3. **Bellini LM, Shea JA.** Mood change and empathy decline persist during 3 years of internal medicine training. Acad Med. 2005;80(2):164–7.
4. ACGME competencies Available at: http://www.acgme.org/outcome/Comp/compFull.asp Accessed January 2010.
5. **Bigby JA, ed.** Cross-Cultural Medicine. Philadelphia, PA: American College of Physicians; 2003. Chapter One: Beyond Culture: Strategies for Caring for Patients from Diverse Racial, Ethnic, and Cultural Groups. pp 20–21
6. **Smedley BD, Stith AY, Nelson AR, eds.** Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care. Washington, DC: National Academy Press; 2002. Chapter 4: Assessing Potential Sources of Racial and Ethnic Disparities in Care: The Clinical Encounter.
7. **Kleinman A, Eisenberg L, Good B.** Culture, illness, and care: clinical lessons from anthropologic and cross-cultural research. Ann Intern Med. 1978;88(2):251–8.
8. **Berlin EA, Fowkes WC.** A teaching framework for cross-cultural health care: application in family practice. West J Med. 1983;139(6):934–8.
9. **Levin S, Like R, Gottlieb J.** ETHNIC: a framework for culturally competent clinical practice. In Appendix: useful clinical interviewing mnemonics. Patient Care. 2000;34(9):188–9.
10. **Betancourt J, Carrillo J, Green A.** Hypertension in multicultural and minority populations: linking communication to compliance. Curr Hypertens Rep. 1999;1(6):482–8.
11. **Pinderhughes E.** Understanding Race, Ethnicity and Power: The Key to Efficacy in Clinical Practice. New York: Free Press; 1989.
12. **Dobbie A, Medrano M, Tysinger J, et al.** The BELIEF instrument: a preclinical teaching tool to elicit patients' health beliefs. Fam Med. 2003;35(5):316–9.
13. **Kumagai AK, Lypson ML.** Beyond cultural competence: critical consciousness, social justice, and multicultural education. Acad Med. 2009;84(6):782–7.
14. **Saha S, Beach MC, Cooper L.** Patient centeredness, cultural competence and healthcare quality. J Nat Med Assoc. 2008;100(11):1275.
15. **Teal CR, Street RL.** Critical elements of culturally competent communication in the medical encounter: a review and model. Soc Sci Med. 2009;68:533–43.
16. **Cooper L, Roter DL.** Patient-provider communication: the effect of race and ethnicity on process and outcomes of healthcare in Unequal Treatment: Confronting Racial and Ethnic Disparities in Healthcare. Smedly BC, Stith AY, Nelson AR, eds. Institute of Medicine. 2002:552–93
17. The Commonwealth Fund 2001 Health Care Quality Survey. [cited 2009; Available at: http://www.commonwealthfund.org/Content/Surveys/2001/2001-Health-Care-Quality-Survey.aspx. Accessed January 2010.
18. **Neeraj KA, Gustafson DH.** Perceived helpfulness of physicians' communication behavior and breast cancer patients' level of trust over time. J Clin Oncol. 2009;24(2):252–5.
19. **Washington D, Bowles J, Saha S, et al.** Transforming clinical practice to eliminate racial and ethnic disparities in health care. JGIM. 2008;23(5):685–91.
20. **Lang F, Floyd MR, Beine KL, et al.** Sequenced questioning to elicit the patient's perspective on illness: effects on information disclosure, patient satisfaction, and time expenditure. Fam Med. 2002;34(5):325–30.
21. **Mookadam F, Arthur HM.** Social support and its relationship to morbidity and mortality after acute myocardial infarction: systematic overview. Arch Intern Med. 2004;164(14):1514–8.
22. **Pennebaker J, O'Hearn R.** Confidants' feedback and traumatic life events. J Trauma Stress. 1995;8(1).
23. **DiClemente C, Prochaska J, Gibertini M.** Self-efficacy and the stages of self-change of smoking. Cogn Ther Res. 1985;9(2).
24. **Haidet P, Paterniti DA.** "Building" a history rather than "taking" one: a perspective on information sharing during the medical interview. Arch Intern Med. 2003;163(10):1134–40.
25. **Dyche L, Swiderski D.** The effect of physician solicitation approaches on ability to identify patient concerns. J Gen Intern Med. 2005;20(3):267–70.
26. **Corbie-Smith G, Thomas SB, St. George DMM.** Distrust, race, and research. Arch Intern Med. 2002;162(21):2458–63.
27. **Gordon HS, Street RL Jr, Sharf FM, et al.** Racial differences in trust and lung cancer patients' perceptions of physician comunication. J Cln Oncol. 2006;24(6):904–9.
28. The Kaiser Family Foundation National Survey of Physicians Part 1: Doctors on Disparities in Medical Care. March 2002 http://www.kff.org/minorityhealth/20020321a-index.cfm Accessed January 2010.
29. **Cooper LA, Roter DL, Johnson RL, et al.** Patient-centered communication, ratings of care, and concordance of patient and physician race. Ann Int Med. 2003;139(11):907–15.
30. **Blanchard J, Nayar S, Lurie N.** Patient-provider and patient-staff racial concordance and perceptions of mistreatment in the health care setting. JGIM. 2007;22(8):1184–9.
31. **Casagrande SS, Gary TL, LaVeist DJ, et al.** Perceived discrimination and adherence to medical care in a racially integrated community. JGIM. 2007;22(3):389–95.
32. **Boulware LE, Cooper LA, Ratner LE, et al.** Race and trust in the health care system. Public Health Rep. 2003;118(4):358–65.
33. **Cooper LA, Beach MC, Johnson RL, et al.** Delving below the surface. Understanding how race and ethnicity influence relationships in health care. JGIM. 2006;21(Suppl 1):S21–7.
34. **Blanchard J, Lurie N.** R-E-S-P-E-C-T: patient reports of disrespect in the health care setting and its impact on care. J Fam Pract. 2004;53(9):721–30.
35. **Corbie-Smith G.** The continuing legacy of the Tuskegee Syphilis Study: considerations for clinical investigation. Am J Med Sci. 1999;317(1):5–8.
36. **Johnson RL, Roter D, Powe NR, et al.** Patient race/ethnicity and quality of patient-physician communication during medical visits. Am J Publ Health. 2004;94(12):2084–90.
37. **Mast MS, Hall JA, Roter DL.** Caring and dominance affect participants' perceptions and behaviors during a virtual medical visit. JGIM. 2008;23(5):523–7.

38. **Green AR, Betancourt JR, Carrillo JE.** Integrating social factors into cross-cultural medical education. Acad Med. 2002;77(3):193–7.
39. **Spiegel D.** Healing words: emotional expression and disease outcome. JAMA. 1999;281(14):1328–9.
40. **Smyth JM, Stone AA, Hurewietz A, Kaell A.** Effects of writing about stressful experiences on symptom reduction in patients with asthma or rheumatoid arthritis: a randomized trial. JAMA. 1999;281(14):1304–9.
41. **Mauksch LB, Dugdale D, Dodson S, Epstein R.** Relationship, communication and efficiency in the medical encounter: creating a clinical model from a literature review. Arch Intern Med. 2008;168(13):1387–95.
42. **Hafferty FW.** Beyond curriculum reform: confronting medicine's hidden curriculum. Acad Med. 1998;74(3):403–7.
43. **Hafferty FW, Frank R.** The hidden curriculum, ethics teaching, and the structure of medical education. Acad Med. 1994;69(11):861–71.
44. **Braddock CH, Eckstrom E, Haidet P.** The "New Revolution" in medical education: fostering professionalism and patient-centered communication in the contemporary environment. JGIM 2004;19(5 part 2):610–11.
45. **Murray-Garcia JL, Garcia J.** The institutional context of multicultural education: what is your institutional curriculum? Acad Med. 2008;83 (7):646–66.
46. **Hojat M, Vergare MJ, Maxwell K, Brainard G, Herrine SK, Isenberg GA, Veloski J, Gonnella JS.** The devil is in the third year: a longitudinal study of erosion of empathy in medical school. Academic Medicine. 2009;84(9):1182–1191.
47. **Lurie N, Yergan J.** Teaching residents to care for vulnerable populations in the out-patient setting. JGIM. 1990;5:527–34.
48. **Bosk CL.** Forgive and Remember: Managing Medical Failure. 2nd ed. Chicago: University of Chicago Press; 2003.
49. **Kohn LT, Corrigan JM, Donaldson MS, eds.** To Err is Human: Building A Safer Health System. Committee on Quality of Health Care in America, Institute of Medicine; 1999.
50. **Branch WT Jr, Kern D, Haidet P, et al.** Teaching the human dimensions of care in clinical settings. JAMA. 2001;286(9):1067–74.
51. **Haidet P, Stein H.** The role of the student-teacher relationship in the formation of physicians. JGIM. 2006;21(S1):S16–S20.
52. **Kern DE, Branch WT Jr, Jackson JL, Brady BW, Feldman MD, Levinson W, Lipkin M Jr.** Teaching the psychosocial aspects of care in the clinical setting: practical recommendations. Acad Med. 2005;80(1):8–20.
53. **Bigby J, Ko LK, Johnson N, David MM, Ferrer B, REACH Boston 2010.** Breast and cervical cancer coalition: a community approach to addressing excess breast and cervical cancer mortality among women of African descent in Boston. Public Health Reports. 2003;118(4):334.
54. **Saitz R, et al**. Alcohol Clinical Training Project, Implementation and evaluation report. supported by the National Institute on Alcohol Abuse and Alcoholism (NIAAA) R25 AA013822, www.alcoholtraining.org, July,2007:25-9.
55. Denial A, Hoppe E, Carlson N. Assessing cultural competence in optometric faculty. Optom Educ. 2006;92-5.
56. Assessing competence in communication and interpersonal skills: The Kalamazoo II Report. Acad Med. 2004;79:495-507.
57. **Parrot S, et al.** Evidence-based office teaching: the five-step microskills model of clinical teaching. Fam Med. 2006;38(3):164–7.
58. **Neher J.** A five-step "microskills" model of clinical teaching. J Am Board Fam Pract. 1992;5(4):419–24.
59. **Mostow C, Principal investigator; Gordon S, Co-principal investigator.** Precepting for Cultural Competence. BCBSM Medical Foundation 2004-5.
60. www.aachonline.org Accessed January 2010.
61. **Campinha-Bacote J.** The process of cultural competence in the delivery of healthcare services: a model of care. J Transcult Nurs. 2002 Jul;13 (3):181-4; 200-1.
62. **Jones C.** Confronting Institutionalized Racism. Phylon 2002 (1960-) 50 (1/2):7-22 Also available at http://www.jstor.org/stable/4149999
63. **Stern DT.** In search of the informal curriculum: when and where professional values are taught. Acad Med. 1998;73(suppl 20):S38–30.
64. **Suchman AL, Williamson PR, Litzelman DK, et al.** The Relationship-centered discovery team: toward an informal curriculum that teaches professionalism: transforming the social environment of a medical school. J Gen Intern Med. 2004;19(5 pt2):501–4.
65. **Snyder L, Leffler C, Ethics and Human Rights Committee, American College of Physicians**. Ethics Manual, 5th ed. Ann Intern Med. 2005;142:560–82.

Addiction
COMMENTARY

Commentary on McCaul *et al.* (2010): Observational studies about average alcohol consumption and health - closing time for a limited evidence base

Positive attitudes about 'moderate' drinking may weaken efforts designed to reduce alcohol misuse and its terrible consequences. For many, it is dogma that alcohol is good for the heart, if not one's overall health. The study by McCaul *et al.* [1] about the relationship between alcohol consumption, mortality and coronary heart disease (CHD) death among older Australians will not change those convictions. Although studies about drinking and mortality are mixed, this study's findings for CHD are generally consistent with those from other studies [2]; and although it is rarely emphasized, this study is also consistent with others in demonstrating that, among drinkers, drinking less alcohol generally results in better health outcomes than drinking more [3]. However, this study suffers from the same limitations that afflict other observational studies about average alcohol consumption and chronic disease outcomes. While these limitations are sometimes overlooked or dismissed, they are worth reviewing in depth and with a frequency matching the feverish publicity heralding new alcohol studies.

While confounding is an important theoretical consideration in any observational study, evidence suggests that confounding is a serious problem in alcohol studies conducted among western populations. First, many traditional CHD risk factors are more prevalent and intense among non-drinkers [4,5], and many analyses attempt to control for these differences statistically. In this paper, smoking was one of the few confounders controlled for in statistical analyses. However, smoking is one CHD risk factor that is typically more prevalent among drinkers, so its inclusion in analytical models, coupled with a failure to control for other confounders, probably led to overestimates of the protective effect of alcohol. Even in more carefully controlled studies, however, residual confounding would probably bias studies in favor of moderate drinkers. Furthermore, those with more risk factors have more possible combinations of risk factors that could be synergistic in terms of CHD risk. To the extent that this synergistic risk is not captured in observational studies, this would also bias studies in favor of moderate drinkers. Finally, because CHD risk factors tend to cluster in certain individuals and populations, it seems plausible that unknown or unmeasured confounders would also be more prevalent among non-drinkers, further biasing studies in favor of moderate drinkers.

In addition to the distribution of traditional cardiac risk factors, moderate average alcohol consumption appears to be a marker of affluence, leisure, education, social advantage, good mental health and having nice teeth (literally) [5]. These psycho-socio-economic markers, many of which are also considered 'non-traditional' cardiac markers are, in turn, major determinants of mortality [6]. Because there is not a plausible causal relationship between, for example, alcohol and higher educational attainment, it seems particularly likely that moderate drinking is merely a reflection of prosperity and wellness, not its genesis.

The second set of limitations—especially important among studies of older people—could be grouped under the heading of selection bias (i.e. when there is a systematic error in the selection or enrollment of study subjects that distorts the relationship between alcohol and outcomes). Of these, the 'sick quitter' bias is discussed most frequently. Those who are ill or frail often stop drinking, thus contaminating non-drinking study groups with unhealthy people who were exposed previously to alcohol [4,7]. In this paper, information about former drinking status was unavailable for women and was not utilized to separate former-drinking from non-drinking men. While most researchers mitigate the sick quitter problem by excluding former drinkers from their analyses altogether, this still biases results against non-drinkers as it selectively removes a frail population whose poor health outcomes would have otherwise accrued to the drinking group in a randomized trial with an intention-to-treat protocol.

In addition, established moderate drinkers enrolled in observational studies are undoubtedly different than non-drinkers who might be randomized to drink in a trial: they self-selected to drink alcohol; they tolerated or enjoyed its effects; they did not die or stop drinking due to health or social problems (alcohol-related or otherwise) prior to the inception of the survey or study cohort; they continued to drink moderately in order to meet enrollment criteria; and they were of sufficient physical and mental capacity to be study respondents. Finally, rates of binge drinking (i.e. drinking at levels that typically result in impairment) may be considerably lower among study participants than the general population [8,9], and many with moderate average consumption also binge drink [10]. Because binge drinking is associated with increased mortality and CHD and a loss of any protective associations with moderate average alcohol consumption [11,12], the associations observed in studies may not be realized in the general population.

The ultimate relevance of alcohol research lies in the domains of public health, clinical care and individual behavior. Many scientists and physicians have implicitly or explicitly framed low-dose alcohol consumption as a

potential therapeutic agent [13], and the public is similarly engaged. To date, however, there has not been a single randomized trial of low-dose alcohol and any mortality outcome. Such trials, which are typically required to evaluate new pharmaceutical agents, should be the standard to which low-dose alcohol is held, particularly as alcohol misuse is a leading public health problem world-wide [14]. In the recent past, numerous observational studies suggested that beta carotene intake was associated with reductions in CHD and cancer, that hormone replacement therapy and vitamin E supplementation were associated with reductions in CHD and dementia and that *Chlamydia* infection was associated with CHD. However, supplementation with hormones, beta-carotene and vitamin E and anti-microbial treatment for *Chlamydia* were found to be ineffective or harmful when subjected to randomized controlled trials [15–19].

While it is possible that low-dose alcohol may be beneficial for some or even many health outcomes, the appeal of possible health panaceas generally, and of alcohol in particular, may have resulted in a reduced scientific standard when considering low-dose alcohol as a potential therapeutic agent. Therefore, in the absence of evidence from randomized trials and weighing the real-world implications of public messages promoting alcohol consumption, we would be best served by aggressively implementing effective policy and clinical interventions to reduce excessive per-occasion alcohol consumption among those who already drink.

Declaration of interests

None.

Keywords Alcohol, cardiovascular disease, moderate drinking, mortality.

TIMOTHY NAIMI
Boston Medical Center—General Internal Medicine, Boston, MA, USA. E-mail: tim.naimi@bmc.org

References

1. McCaul K. A., Almeida O. P., Hankey G. J., Jamrozik K., Byles J. E., Flicker L. Alcohol use and mortality in older men and women. *Addiction* 2010; **105**: 1391–400.
2. Corrao G., Rubbiati L., Bagnardi V. *et al*. Alcohol and coronary heart disease: a meta-analysis. *Addiction* 2000; **95**: 1505–23.
3. Kloner R. A., Shereif R. H. To drink or not to drink? That is the question. *Circulation* 2007; **116**: 1306–17.
4. Wannamethee G., Shaper A. G. Men who do not drink: a report from the British Regional Heart Study. *Int J Epidemiol* 1988; **17**: 307–16.
5. Naimi T. S., Brown D. W., Brewer R. D. *et al*. Cardiovascular risk factors and confounders among nondrinking and moderate-drinking U.S. adults. *Am J Prev Med* 2005; **28**: 369–73.
6. Marmot M. Social determinants of health inequalities. *Lancet* 2005; **365**: 1005–6.
7. Fillmore K. M., Kerr W. C., Stockwell T. *et al*. Moderate alcohol use and reduced mortality risk: systematic error in prospective studies. *Addict Res Theory* 2006; **14**: 101–32.
8. Jousilahti P., Veikko S., Kuulasmaa K. *et al*. Total and cause-specific mortality among participants and non-participants of population-based health surveys: a comprehensive followup of 54,372 Finnish men and women. *J Epidemiol Commun Health* 2005; **59**: 310–15.
9. Rosengren A., Wilhelmsen L., Berglund G., Elmfeldt D. Non-participants in a general population study of men, with special reference to social and alcoholic problems. *Acta Med Scand* 1987; **221**: 243–51.
10. Naimi T. S., Brewer R. D., Mokdad A. H. *et al*. Binge drinking among U.S. adults. *JAMA* 2003; **289**: 70–5.
11. McElduff P., Dobson A. J. How much alcohol and how often? Population based case-control study of alcohol consumption and risk of a major coronary event. *BMJ* 1997; **314**: 1159–64.
12. Roerecke M., Rehm J. Irregular heavy drinking occasions and risk of ischemic heart disease: a systematic review and meta-analysis. *Am J Epidemiol* 2010; DO1: 10.1093/aje/kwp1451.
13. Goldberg I. J. To drink or not to drink? *N Eng J Med* 2003; **348**: 163–4.
14. Ezzati M., Lopez A. D., Rodgers A. *et al*. Selected major risk factors and global and regional burden of disease. *Lancet* 2002; **360**: 1347–60.
15. Anderson J. L. Infection, antibiotics, and atherothrombosis—end of the road or new beginnings? *N Eng J Med* 2005; **352**: 1706–9.
16. Blacker D. Mild cognitive impairment—no benefit from vitamin E, little from donepezil. *N Eng J Med* 2005; **352**: 2439–41.
17. Greenberg E. R. Antioxidant vitamins, cancer, and cardiovascular disease. *N Eng J Med* 1996; **334**: 1189–90.
18. Women's Health Initiative Investigators. 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.
19. Yaffe K. Hormone therapy and the brain—deja vu all over again? *JAMA* 2003; **289**: 2717–19.

The Intensity of Binge Alcohol Consumption Among U.S. Adults

Timothy S. Naimi, MD, MPH, David E. Nelson, MD, MPH, Robert D. Brewer, MD, MSPH

Background: Binge drinking (consuming five or more drinks during a drinking occasion) is responsible for more than half of the 79,000 annual deaths due to excessive drinking in the U.S. Although studies show a strong dose–response relationship between the intensity of binge drinking (i.e., the number of drinks consumed per binge episode) and adverse outcomes, there are no population-based studies assessing this measure.

Purpose: This study aims to analyze population-based data from a module of questions on binge drinking among U.S. adults to assess the number of drinks consumed by binge drinkers and the associated independent risk factors for consuming more drinks.

Methods: Data were analyzed from 14,143 adult binge drinkers who responded to the Behavioral Risk Factor Surveillance System binge drinking module in 2003 and 2004. Total drinks were calculated by summing the total number of beer, wine, and liquor-containing drinks consumed during a respondents' most recent binge drinking episode.

Results: Binge drinkers consumed an average of 8.0 drinks (median 6) during their most recent binge drinking episode; 70.0% of binge drinkers consumed six or more drinks, and 38.4% consumed eight or more drinks. Men consumed more drinks during their last binge episode than women (M=8.3 vs 7.0, median=7 vs 6), and those aged 18–34 years consumed more drinks than those aged >34 years for both men and women. Independent risk factors for consuming eight or more drinks included being male; being aged <35 years; being other than white race/ethnicity; having less education; not being married; binge drinking three or more times in the past 30 days; and drinking mostly beer.

Conclusions: Most adult binge drinkers drink in excess of the five-drink threshold defining this risky behavior. The intensity of binge drinking should be monitored regularly by health agencies to improve surveillance and to better assess the impact of interventions designed to reduce binge drinking and its consequences.

(Am J Prev Med 2010;38(2):201–207) Published by Elsevier Inc. on behalf of American Journal of Preventive Medicine

Introduction

Excessive alcohol consumption is a leading preventable cause of death[1] that causes approximately 79,000 deaths annually in the U.S., and it shortens the lives of those who die by approximately 30 years.[2] Binge drinking, often defined as the consumption of five or more drinks on an occasion, typically results in acute impairment and is responsible for more than half of these deaths.[2,3] Binge drinking is also a leading risk factor for a variety of health and social outcomes, such as unintentional injuries, violence, unintended pregnancy, and cardiovascular disease.[4–8]

Studies have demonstrated that the risk of alcohol-related harms increases with the intensity of binge drinking—that is, with an increased number of drinks consumed per binge episode and with higher blood alcohol concentrations.[9–12] However, we are unaware of prior studies that have characterized the number of drinks consumed per binge episode on a population basis. Such information is crucial for characterizing the risks associated with binge drinking and for evaluating the impact of strategies to prevent this behavior. This study used population-based data from a module of questions on binge drinking among U.S. adults to assess the number of drinks consumed by binge drinkers and independent risk factors for consuming more

From the Alcohol Team, Emerging Investigations and Analytic Methods Branch, Division of Adult and Community Health, National Center for Chronic Disease Prevention and Health Promotion, CDC, Atlanta, Georgia

Address correspondence and reprint requests to: Timothy S. Naimi, MD, MPH, Boston Medical Center, GIM, 801 Massachusetts Avenue, Crosstown Building, 2nd Floor, Boston MA 02118. E-mail: tbn7@cdc.gov.

0749-3797/00/$17.00

doi: 10.1016/j.amepre.2009.09.039

drinks during a person's most recent binge drinking episode.

Methods

Data for this study came from the CDC's Behavioral Risk Factor Surveillance System (BRFSS). Extensive detail about the BRFSS and its methods are available at www.cdc.gov/brfss/. In brief, the BRFSS includes state-based random-digit-dial telephone surveys of people aged ≥18 years, which are conducted monthly in all states, the District of Columbia, and some territories; survey instruments contain questions on a variety of health risk measures, including alcohol consumption. Data are weighted to be representative of states or other jurisdictions.

The current study was restricted to binge drinkers identified from the BRFSS core survey (i.e., the portion of the survey asked of all respondents in all states and territories) in 2003 and 2004. A binge drinker was defined as someone who consumed alcohol in the past 30 days who gave a nonzero response to the following question: *Considering all types of alcoholic beverages, how many times during the past 30 days did you have five or more drinks on an occasion?*

In 2003 and 2004, the CDC offered states an optional module of six additional questions for people who reported one or more occasions of binge drinking in the past 30 days; all the questions in the module pertained to a respondent's most recent binge drinking episode. These questions asked binge drinkers about the number and type of alcohol-containing beverages (beer, wine, or liquor) consumed during their most recent binge drinking episode; the physical location of their last binge episode; and whether the respondent drove during or within 2 hours of their binge drinking episode. The intensity of binge drinking—that is, the total number of drinks consumed during the most recent binge episode—was calculated by summing the number of beer, wine, and liquor-containing drinks.

Analyses were limited to 18 states that used the BRFSS Binge Drinking optional module in 2003–2004 (13 states in 2003 and 14 states in 2004). States using the module in both years were California, Maine, Michigan, Minnesota, Montana, Nevada, New Hampshire, Wisconsin, and Wyoming; states using this module in 2003 only were Nebraska, North Carolina, Pennsylvania, and South Dakota; and states using the module in 2004 only were Delaware, Idaho, New Mexico, North Dakota, and Virginia. Data were weighted for age, gender, and race to be representative of the states and years for which data were analyzed; state weights were divided by 2 for those states with data from both years. The median response rate to the BRFSS survey among these states across both years was 55.0% and included 121,172 respondents, including 16,496 people who reported one or more episodes of binge drinking in the past month. The weighted prevalence of binge drinking was 16.3%, which was approximately 1 percentage point higher than for the U.S. during 2003–2004. After excluding those with missing or incomplete information from the binge drinking module, data from 14,143 respondents were analyzed.

Data analyses were conducted using SAS, version 9.0, and SUDAAN, version 9.0. Analyses were conducted for demographic characteristics, social characteristics, and alcohol-related covariates. Demographics included age groups, gender, race/ethnicity, education level, income level, marital status, and employment status. Alcohol variables included number of binge episodes in the past 30 days, predominant beverage type consumed during the most recent binge episode, location of the binge drinking episode, and whether or not the respondent drove a motor vehicle during or within 2 hours of his or her most recent binge drinking episode.

We used logistic regression analysis to further explore the relationship between high-intensity binge drinking (i.e., consuming eight or more drinks per binge episode) and those demographic and alcohol variables that had a significant effect on the mean number of binge drinks consumed in bivariate analyses. Because of their important associations with alcohol consumption measures in other studies, more detailed information about age, race/ethnicity, and education were presented in Table 1. However, because of similarities among certain groups and sample size considerations, those categories were collapsed for regression analysis. Age categories were collapsed to those aged 18–34 years (mean number of drinks=8.7; 95% CI=8.5, 8.9); those aged 35–54 years (M=7.4 drinks, 95% CI=7.2, 7.6); and those aged ≥55 years (M=6.7 drinks; 95% CI=6.4, 7.1). Race/ethnicity categories were collapsed to white, non-Hispanic (M=7.7 drinks; 95% CI=7.6, 7.9) and other than white non-Hispanic (M=8.8 drinks; 95% CI=8.4, 9.2). Education categories were collapsed to being a high school graduate or less (M=8.7 drinks; 95% CI=8.4, 9.0) and having at least some college education (M=7.5 drinks; 95% CI=7.4, 7.7).

Results

The study population was 75.1% male, with 52.7% aged ≤34 years, 59.4% with at least some college education, and 72.8% of white non-Hispanic race/ethnicity. Overall, binge drinkers consumed an average of 8.0 (median=6) drinks during their most recent binge drinking episode (Table 1). Groups that consumed a significantly higher number of binge drinks included those who were male; aged 18–34 years; belonged to racial/ethnic groups other than white non-Hispanic (data not shown); had no college education (data not shown); earning <$50,000 per year; not married; those with five or more binge episodes in past 30 days; and those who consumed predominantly beer. Men consumed more drinks during their most recent binge drinking episode compared with women (M=8.3 vs 7.0, median=7 vs 6); however, stratification by gender did not significantly change the relative distri-

Table 1. Total drinks consumed in most recent binge drinking[a] episode, by sociodemographic characteristics and patterns of alcohol consumption

	No.	All[b]		Men[b] (*n*=9655)	Women[b] (*n*=4488)
Characteristic		**M (95% CI)**	**Median**	**M (95% CI)**	**M (95% CI)**
All	14,143	8.0 (7.9, 8.2)	6	8.3 (8.2, 8.5)	7.0 (6.8, 7.2)
Age (years)					
18–24	2,079	9.5 (9.1, 9.8)	8	10.1 (9.6, 10.5)	8.1 (7.5, 8.7)
25–34	3,626	8.0 (7.7, 8.3)	7	8.4 (8.0, 8.8)	6.8 (6.6, 7.0)
35–44	3,672	7.4 (7.2, 7.6)	6	7.7 (7.4, 7.9)	6.5 (6.3, 6.8)
45–54	2,856	7.4 (6.9, 7.8)	6	7.8 (7.2, 8.2)	6.1 (5.8, 6.3)
≥55	1,910	6.7 (6.4, 7.1)	6	6.9 (6.5, 7.3)	6.1 (5.4, 6.8)
Race					
White, non-Hispanic	12,110	7.7 (7.6, 7.9)	6	8.1 (7.9, 8.2)	6.9 (6.7, 7.1)
Hispanic	881	9.0 (8.4, 9.5)	7	9.2 (8.6, 9.8)	7.5 (6.5, 8.4)
Black	376	8.2 (7.5, 9.0)	6	8.4 (7.5, 9.4)	7.8 (6.7, 8.9)
Other	706	8.6 (7.7, 9.4)	7	9.0 (8.0, 10.0)	7.1 (6.4, 7.8)
Education					
<High school graduate	904	8.7 (8.4, 9.0)	7	10.0 (9.0, 11.0)	7.0 (6.3, 7.8)
High school graduate	4,476	9.5 (8.6, 10.3)	7	8.7 (8.4, 9.0)	7.7 (7.1, 8.2)
Some college	4,337	8.0 (7.8, 8.2)	7	8.4 (8.1, 8.7)	7.0 (6.8, 7.3)
College graduate	4,416	7.0 (6.9, 7.2)	6	7.3 (7.1, 7.5)	6.3 (6.1, 6.6)
Income ($)					
0–49,000	7,471	8.3 (8.1, 8.5)	7	8.7 (8.4, 8.9)	7.2 (6.9, 7.5)
≥50,000	5,784	7.6 (7.4, 7.8)	6	7.9 (7.6, 8.1)	6.6 (6.3, 7.0)
Marital status					
Married	6,827	7.4 (7.2, 7.6)	6	7.6 (7.4, 8.0)	6.3 (6.1, 6.5)
Not married[c]	7,301	8.6 (8.4, 8.8)	7	9.0 (8.8, 9.3)	7.5 (7.2, 7.8)
Employment					
Employed	11,243	8.0 (7.8, 8.2)	6	8.3 (8.1, 8.5)	7.1 (6.8, 7.3)
Not employed[d]	2,889	8.1 (7.8, 8.4)	7	8.6 (8.2, 9.1)	6.9 (6.6, 7.2)
BMI					
<30	11,105	7.9 (7.8, 8.1)	6	8.3 (8.1, 8.5)	6.9 (6.7, 7.2)
≥30	3,038	8.3 (7.9, 8.8)	7	8.6 (8.1, 9.1)	7.4 (6.8, 8.0)
Binge episodes, past 30 days					
1–2	8,291	7.2 (7.0, 7.3)	6	7.5 (7.3, 7.7)	6.4 (6.3, 6.6)
3–4	2,637	8.2 (7.9, 8.6)	7	8.5 (8.1, 8.9)	7.4 (6.9, 8.0)
≥5	3,215	9.8 (9.4, 10.2)	8	10.0 (9.5, 10.4)	8.8 (8.0, 9.6)
Predominant beverage type[e]					
Beer	10,764	8.3 (8.1, 8.5)	7	8.6 (8.4, 8.8)	7.3 (7.0, 7.6)
Wine or liquor	3,379	7.1 (6.8, 7.3)	6	7.4 (7.0, 7.7)	6.6 (6.3, 7.0)

(continued on next page)

Table 1. Total drinks consumed in most recent binge drinking[a] episode, by sociodemographic characteristics and patterns of alcohol consumption *(continued)*

	No.	All[b]		Men[b] (*n*=9655)	Women[b] (*n*=4488)
Characteristic		**M (95% CI)**	**Median**	**M (95% CI)**	**M (95% CI)**
Drinking location[f]					
Private residence	7,609	8.0 (7.8, 8.2)	6	8.3 (8.0, 8.6)	7.1 (6.8, 7.4)
Licensed outlet	5,109	7.9 (7.6, 8.1)	6	8.3 (8.0, 8.6)	6.9 (6.6, 7.3)
Other	1,389	8.4 (8.0, 8.9)	7	8.9 (8.3, 9.5)	6.9 (6.5, 7.4)
Drove motor vehicle[g]					
Yes	1,846	8.3 (7.9, 8.8)	7	8.5 (8.0, 9.0)	7.3 (6.6, 8.0)
No	12,232	8.0 (7.8, 8.1)	6	8.3 (8.1, 8.5)	7.0 (6.8, 7.2)

[a]Binge drinking was defined as consuming five or more drinks on at least one occasion in the past 30 days. Respondent information pertained to their most recent episode of binge drinking. Strata of respondent numbers may not add up to 14,143 because of nonresponse to selected questions (e.g., income).
[b]All results were weighted by age, gender, and race/ethnicity to be representative of people residing in the states and years for which data were analyzed.
[c]Not married included those who were never married, divorced, separated, or widowed.
[d]Not employed people included those who were unemployed, home makers, students, retired, or unable to work.
[e]Predominant beverage type referred to whether the respondent drank mostly beer, liquor, or wine or flavored beverages during their most recent binge drinking episode.
[f]Private residences included the drinker's own home or another person's home; licensed outlets included bars, clubs, or restaurants; others included all other locations.
[g]Driving after binge drinking refers to those who reported driving "during or within 2 hours" of their most recent binge drinking episode.

bution of drinks per binge episode for most characteristics. Those consuming alcohol in establishments licensed to sell alcohol (bars, clubs, or restaurants) consumed a similar amount (M=8.0 drinks, median=6.0 drinks) compared to those who drank in their own or another person's private residence (M=8.0 drinks, median=6.0 drinks).

During their most recent binge drinking episode, 70.0% of drinkers consumed six or more drinks, 38.4% consumed eight or more drinks, and 16.9% consumed 11 or more drinks (Table 2). Men and binge drinkers aged 18–34 years were more likely to consume larger numbers of drinks per binge episode than were women and people aged ≥35 years, respectively. For example, more than half (51.1%) of the male binge drinkers aged 18–34 years reported consuming eight or more drinks during their most recent binge drinking episode, as compared to 31.4% of women in this age group. Even so, a substantial proportion of male and female binge drinkers aged ≥55 years also reported consuming six or more drinks during their most recent binge episode (58.1% of men and 38.2% of women, although numbers are small).

In logistic regression analysis, independent risk factors for high-intensity binge drinking (i.e., consuming eight or more drinks per episode) included being male; being of younger age; being a member of a racial/ethnic group other than white non-Hispanic; having no college education; not being married; binge drinking on three or more occasions in the past 30 days; and drinking mostly beer (Table 3). Income level was not associated with consuming more drinks during the last binge episode.

Discussion

To our knowledge, this is the first study to examine the intensity of binge drinking, that is, the total number of drinks consumed during a discrete binge drinking episode. It was found that adult binge drinkers in the U.S. consumed an average of eight drinks per binge episode, and that 70% of binge drinkers consumed six or more drinks, including almost 40% who consumed eight or more drinks. Although drinking just five drinks on one occasion is associated with adverse outcomes in epidemiologic studies and typically results in blood alcohol concentrations of 0.08 mg/dL or higher,[3,10,13,14] the risk of adverse outcomes (e.g., unintentional injuries) increases with the number of drinks consumed or with progressively higher blood alcohol concentrations.[9,11,12] As such, these findings illuminate an important dimension of binge alcohol consumption, and emphasize the need to reduce the intensity of binge drinking among U.S. adults as well as its prevalence and frequency.

These findings are concordant with other results[15] demonstrating that men and younger people are more

Table 2. Distribution of number of drinks consumed during most recent binge drinking episode,[a] by gender and age

Characteristic	No. of drinks consumed by binge drinkers,[b] % (95% CI)			
	5 drinks	6–7 drinks	8–10 drinks	≥11 drinks
All binge drinkers	30.0 (28.5, 31.4)	31.7 (30.2, 33.2)	21.5 (20.1, 22.9)	16.9 (15.5, 18.2)
Men by age (years)				
All	26.0 (24.3, 27.7)	31.5 (29.7, 33.3)	23.2 (21.5, 25.0)	19.3 (17.1, 21.1)
18–34	19.2 (16.8, 21.8)	29.7 (27.0, 32.6)	24.8 (22.2, 27.6)	26.3 (23.6, 29.2)
35–54	30.3 (27.8, 32.9)	33.6 (31.0, 36.2)	22.4 (20.0, 25.0)	13.7 (11.7, 16.0)
≥55	42.0 (37.1, 47.0)	32.3 (27.7, 37.2)	18.5 (14.1, 23.9)	7.3 (4.9, 10.8)
Women by age (years)				
All	42.0 (39.2, 44.7)	32.4 (29.9, 35.0)	16.3 (14.3, 18.5)	9.4 (7.8, 11.2)
18–34	35.7 (32.1, 39.6)	32.8 (29.4, 36.5)	19.3 (16.4, 22.5)	12.1 (9.7, 15.1)
35–54	49.1 (45.0, 53.3)	32.8 (29.0, 36.9)	12.5 (10.0, 15.4)	5.6 (4.0, 7.6)
≥55	61.8 (50.3, 72.0)	26.2 (17.4, 37.4)	8.6[c]	3.4[c]

[a]Binge drinking was defined as consuming five or more drinks on at least one occasion in the past 30 days; respondent information pertained to their most recent episode of binge drinking.
[b]The sum of all rows may not equal 100.0% as a result of rounding error.
[c]CIs were not reported because there were less than 50 respondents.

likely to binge drink compared with women and older people. However, it was interesting that non-Hispanic whites and those with higher educational attainment consumed fewer drinks per binge drinking episode compared with people with lower educational levels who were of a race/ethnicity other than white, in view of the fact other studies[16] have shown that the prevalence of binge drinking does not differ dramatically based on educational levels and that the prevalence of binge drinking among white non-Hispanic people is comparable to that of white Hispanics and higher than that for blacks. These inconsistencies emphasize the importance of assessing multiple measures of binge drinking, including intensity, to obtain a more complete picture of binge drinking behavior.

In terms of other findings with relevance to health policy, it was surprising that people binge drinking in establishments licensed to sell alcohol (bars, clubs, and restaurants) consumed as many drinks as those drinking in private residences. Because state laws generally prohibit the sale of alcohol to intoxicated people, our findings underscore the need for better enforcement efforts.[17] However, these results are not surprising in light of studies demonstrating that most pseudo-intoxicated patrons are served alcohol, including approximately half of the instances when the server appears to notice that the patron is intoxicated.[18–20] In addition, the current study found that those drinking mostly beer consumed more drinks compared with those drinking predominantly wine or distilled spirits. This is consistent with the fact that beer accounts for most binge drinks consumed by adults, and that beer is generally taxed, distributed, and marketed more permissively than other beverage types.[21]

This study is subject to several limitations that make it likely that our estimates of the number of drinks consumed per binge episode were conservative. Survey data were from self-report, and survey respondents may under-report how much they drink.[22] Excessive drinking may also be associated with nonresponse,[23,24] and therefore the response rate may have resulted in a conservative estimate of the number of drinks consumed per binge episode. Furthermore, respondents who were unable to recall how many drinks they consumed were excluded from our analyses, and it is possible that these individuals were more likely to have consumed more drinks compared with those who were able to recall the number of drinks they consumed.

Although this was a study that included data from multiple states, these data may not be representative of the entire country, and the prevalence of binge drinking was slightly higher in the study states compared to the U.S. general population. This study examined those who reported binge drinking at least once in the past month, and the number of drinks consumed would have likely been somewhat different if it had included those who binge drank more or less frequently (e.g., those who consumed five or more drinks in the past 2 weeks). Subsequent to this study, the definition of binge drinking has been modified to drinking five or more drinks for men and four or more drinks for women.[3] Were those thresholds used to define binge drinking in this study, women consuming only four drinks would have been included in the study population and their mean number of drinks consumed would have been lower. Finally, although standard drinks sizes contain similar

Table 3. AORs of consuming eight or more drinks during most recent binge drinking episode[a]

Characteristic	Percentage consuming ≥8 drinks	AOR (95% CI) of consuming ≥8 drinks[b]
All	38.4	
Gender		
Men	42.5	1.82 (1.54, 2.17)
Women	25.6	1.0 (ref)
Age (years)		
18–34	45.5	2.27 (1.71, 3.03)
35–54	32.1	1.45 (1.10, 1.93)
≥55	23.5	1.0 (ref)
Race/ethnicity		
Other than white, non-Hispanic	46.3	1.45 (1.18, 1.79)
White, non-Hispanic	35.4	1.0 (ref)
Education level		
≤High school	44.3	1.32 (1.12, 1.56)
≥Some college	34.2	1.0 (ref)
Income level ($)		
0–49,000	42.0	1.01 (0.85, 1.20)
≥50,000	33.3	1.0 (ref)
Marital status		
Not married[c]	45.2	1.44 (1.21, 1.70)
Married	30.7	1.0 (ref)
Binge episodes, past 30 days		
≥5	56.3	2.70 (2.23, 3.27)
3–4	42.8	1.58 (1.30, 1.93)
1–2	29.0	1.0 (ref)
Predominant beverage type[d]		
Mostly beer	42.8	1.81 (1.49, 2.19)
Mostly wine, liquor	24.4	1.0 (ref)

[a]Binge drinking was defined as consuming five or more drinks on at least one occasion in the past 30 days; respondent information pertained to their most recent episode of binge drinking.
[b]OR was adjusted for all covariates listed in the table.
[c]Not married included those who were never married, divorced, separated, or widowed.
[d]Predominant beverage type referred to whether the respondent drank mostly beer, liquor, or wine or flavored beverages during their most recent binge drinking episode.

amounts of ethanol, the actual ethanol consumed per drink may vary by beverage type, drinking location, gender, or other factors.[25]

In light of the high intensity of binge drinking found in the current study, and because subgroups with higher numbers of drinks consumed per binge episode are not necessarily those with the highest prevalence of binge drinking, it is recommended that there be routine monitoring of the number of drinks per binge drinking episode in addition to the prevalence and frequency of this behavior. In addition, public health research is recommended to specifically evaluate the impact of effective policies to reduce binge drinking (e.g., increased alcohol excise taxes) on the number of drinks consumed per binge drinking episode.

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. Mokdad AH, Stroup D, Marks JS, Gerberding J. Actual causes of death in the U.S., 2000. JAMA 2004;291:1238–45.
2. CDC. Alcohol-related disease impact (ARDI). www.cdc.gov/alcohol/ardi.htm.
3. National Institutes of Alcohol Abuse and Alcoholism. NIAAA Council approves binge drinking definition. www.niaaa.nih.gov/publications/Newsletter/winter2004/Newsletter_Number3.htm#council.
4. Chikritzhs TN, Jonas HA, Stockwell TR, et al. Mortality and life-years lost due to alcohol: a comparison of acute and chronic causes. Med J Aust 2001;174:281–4.
5. CDC. Alcohol involvement in fatal motor vehicle crashes, U.S., 1990–2000. MMWR Morb Mortal Wkly Rep 2001;50:1064–5.
6. Naimi TS, Lipscomb LE, Brewer RD, Colley Gilbert B. Binge drinking in the preconception period and the risk of unintended pregnancy: implications for women and their children. Pediatrics 2003;111:1136–41.
7. National Institutes of Alcohol Abuse and Alcoholism. Tenth special report to the U.S. Congress on alcohol and health. Bethesda MD: NIH, 2000.
8. Brewer RD, Swahn MH. Binge drinking and violence. JAMA 2005;294:616–8.
9. Vinson DC, MacLure M, Reidinger C, Smith GS. A population-based case-crossover and case–control study of alcohol and the risk of injury. J Stud Alcohol 2003;64:358–66.
10. Wechsler H, Nelson TF. Relationship between level of consumption and harms in assessing drink cut-points for alcohol research: commentary on "Many college freshmen drink at levels far beyond the binge threshold" by White et al. Alcohol Clin Exp Res 2006;30:922–7.
11. Weitzman ER, Nelson TF. College student binge drinking and the "prevention paradox": implications for prevention and harm reduction. J Drug Educ 2004;34:247–66.

12. Zador PL, Krawchuk SA, Voas RB. Alcohol-related relative risk of driver fatalities and driver involvement in fatal crashes in relation to driver age and gender: an update using 1996 data. J Stud Alcohol 2000;61:387–95.
13. Naimi TS, Brewer RD, Mokdad A, et al. Letter response. JAMA 2003;289:1636.
14. Wechsler H, Austin SB. Binge drinking: the five/four measure. J Stud Alcohol 1998;59:122–4.
15. Naimi TS, Brewer RD, Mokdad AH, et al. Binge drinking among U.S. adults. JAMA 2003;289:70–5.
16. Cremeens J, Nelson DE, Naimi TS, et al. Sociodemographic differences in binge drinking among adults—14 states, 2004. MMWR Morb Mortal Wkly Rep 2009;58: 301–4.
17. Babor TF, Caetano R, Casswell S, et al. Alcohol: no ordinary commodity: research and public policy. New York NY: Oxford University Press, 2003.
18. Freisthler B, Gruenewald PJ, Treno AJ, et al. Evaluating alcohol access and the alcohol environment in neighborhood areas. Alcohol Clin Exp Res 2003;27:477–84.
19. Lenk KM, Toomey TL, Erickson DJ. Propensity of alcohol establishments to sell to obviously intoxicated patrons. Alcohol Clin Exp Res 2006;30:1194–9.
20. Toomey TL, Wagenaar AC, Erickson DJ, et al. Illegal alcohol sales to obviously intoxicated patrons at licensed establishments. Alcohol Clin Exp Res 2004;28:769–74.
21. Naimi TS, Brewer RD, Miller JW, et al. What do binge drinkers drink? Implications for alcohol control policy. Am J Prev Med 2007;33:188–93.
22. Dawson DA. Methodological issues in measuring alcohol use. Alcohol Res Health 2003;27:18–29.
23. Midanik LT. The validity of self-reported alcohol consumption and alcohol problems: a literature review. Br J Addict 1988;77:357–82.
24. Smith PF, Remington PL, Williamson DF, et al. A comparison of alcohol sales data with survey data on self-reported alcohol use in 21 states. Am J Public Health 1990;80:309–12.
25. Kerr WC, Patterson D, Koenan MA, et al. Alcohol content variation of bar and restaurant drinks in Northern California. Alcohol Clin Exp Res 2008;32:1623–9.

Embracing a Health Services Research Perspective on Personal Health Records: Lessons Learned from the VA My HealtheVet System

Kim M. Nazi, PhD(c)[1], Timothy P. Hogan, PhD[2,3,4], Todd H. Wagner, PhD[5,6], D. Keith McInnes, ScD, MS[7,8,9], Bridget M. Smith, PhD[2,3,4,10], David Haggstrom, MD, MAS[11,12,13,14], Neale R. Chumbler, PhD[12,15,16], Allen L. Gifford, MD[7,8,9,17], Kathleen G. Charters, PhD, RN[18], Jason J. Saleem, PhD[11,12,13,19], Kenneth R. Weingardt, PhD[5,6,20,21], Linda F. Fischetti, RN, MS[22], and Frances M. Weaver, PhD[2,3,4,10,23]

[1]Veterans and Consumers Health Informatics Office, Office of Health Information, Veterans Health Administration, US Department of Veterans Affairs, Washington, DC, USA; [2]Health Services Research and Development Center for Management of Complex Chronic Care, US Department of Veterans Affairs, Hines, IL, USA; [3]Spinal Cord Injury Quality Enhancement Research Initiative (QUERI), US Department of Veterans Affairs, Hines, IL, USA; [4]Edward J. Hines, Jr. Veterans Affairs Hospital, Veterans Health Administration, US Department of Veterans Affairs, Hines, IL, USA; [5]Veterans Affairs Palo Alto Healthcare System, Veterans Health Administration, US Department of Veterans Affairs, Palo Alto, CA, USA; [6]Stanford University, Stanford, CA, USA; [7]HIV/Hepatitis Quality Enhancement Research Initiative (QUERI), US Department of Veterans Affairs, Bedford, MA, USA; [8]Health Services Research and Development Center for Health Quality Outcomes and Economic Research, US Department of Veterans Affairs, Bedford, MA, USA; [9]Edith Nourse Rogers Memorial Veterans Hospital, Veterans Health Administration, US Department of Veterans Affairs, Bedford, MA, USA; [10]Institute for Healthcare Studies, Northwestern University, Chicago, IL, USA; [11]Health Services Research and Development Center on Implementing Evidence-Based Practice, US Department of Veterans Affairs, Indianapolis, IN, USA; [12]Richard L. Roudebush VA Medical Center, Veterans Health Administration, Indianapolis, IN, USA; [13]University Center for Health Services and Outcomes Research, Regenstrief Institute, Indianapolis, IN, USA; [14]Division of General Internal Medicine and Geriatrics, Indiana University Department of Medicine, Indianapolis, IN, USA; [15]Stroke Quality Enhancement Research Initiative (QUERI), US Department of Veterans Affairs, Indianapolis, IN, USA; [16]Department of Sociology, Indiana University School of Liberal Arts, Indiana University-Purdue University Indianapolis (IUPUI), Indianapolis, IN, USA; [17]Boston University School of Public Health, Boston University, Boston, MA, USA; [18]Creative Computing Solutions, Inc., Rockville, MD, USA; [19]Department of Electrical and Computer Engineering, Indiana University-Purdue University Indianapolis (IUPUI), Indianapolis, IN, USA; [20]Substance Use Disorders Quality Enhancement Research Initiative (QUERI), US Department of Veterans Affairs, Houston, TX, USA; [21]Michael E. DeBakey VA Medical Center, Veterans Health Administration, Houston, TX, USA; [22]Chief Health Informatics Office, Office of Health Information, Veterans Health Administration, US Department of Veterans Affairs, Washington, DC, USA; [23]Department of Neurology, Northwestern University, Chicago, IL, USA.

BACKGROUND: Personal health records (PHRs) are designed to help people manage information about their health. Over the past decade, there has been a proliferation of PHRs, but research regarding their effects on clinical, behavioral, and financial outcomes remains limited. The potential for PHRs to facilitate patient-centered care and health system transformation underscores the importance of embracing a broader perspective on PHR research.

OBJECTIVE: Drawing from the experiences of VA staff to evaluate the My HealtheVet (MHV) PHR, this article advocates for a health services research perspective on the study of PHR systems.

METHODS: We describe an organizing framework and research agenda, and offer insights that have emerged from our ongoing efforts regarding the design of PHR-related studies, the need to address PHR data ownership and consent, and the promotion of effective PHR research collaborations.

CONCLUSION: These lessons are applicable to other PHR systems and the conduct of PHR research across different organizational contexts.

KEY WORDS: personal health records; health information; health services research; PHR.
J Gen Intern Med 25(Suppl 1):62–7
DOI: 10.1007/s11606-009-1114-6

INTRODUCTION

Recent advances in information and communication technologies have enabled the development of comprehensive tools intended to support greater consumer participation in their healthcare.[1–3] The personal health record (PHR) is one such tool that has potential to dramatically shape the contemporary healtheare landscape. Whereas electronic health records (EHRs) and systems (EHR-S) are collections of health information that are managed by healthcare providers, PHRs are designed to address the health information needs of consumers. Although there is variability in functionality across

systems,[4–11] most PHRs share a basic goal: "to give patients better access to their own healthcare data and enable them to be stewards of their own information."[12]

According to the American Health Information Management Association (AHIMA), the PHR is "an electronic, lifelong resource of health information needed by individuals to make health decisions. Individuals own and manage the information in the PHR, which comes from healthcare providers and the individual. The PHR is maintained in a secure and private environment, with the individual determining rights of access. The PHR does not replace the legal record of any provider."[13] Currently there are more than 200 PHR systems available.[14] Many early PHRs were "static repositories,"[15] but more recently there has been a shift towards web-based PHRs that are integrated with or "tethered" to an EHR-S.[11] Tethered PHRs can bring together data created and stored by the individual with that from the EHR, thus offering a range of functionality.[11,16,17]

The enthusiasm surrounding the development of PHR systems can be attributed to the anticipated value that PHRs hold for consumers, healthcare providers, financers, and other stakeholders;[5,11,12,14,18] however, such perceived benefits extend beyond what is currently known about their use and effects. The perspectives of healthcare providers suggest both excitement over the potential benefits of PHRs and concerns surrounding their impact.[4,5,12,19–22] Trends in consumer survey research reflect limited access to electronic PHRs but suggest growing interest in using them.[23–26] Despite this expressed interest, it remains difficult to anticipate the manner in which different communities of users may ultimately choose to adopt PHRs.

Initiatives undertaken to inform the course of PHR development[27,28] and to articulate important architectural and policy recommendations[29] represent a foundational response to persistent calls for more substantive PHR research;[30] yet, the lack of research demonstrating the value of PHRs to stakeholders poses a threat to their long-term viability and sustainability.[16] Research regarding the effects of PHR use on patient and provider experiences, behavior, costs, and clinical outcomes remains underdeveloped.[11,16] Studies that move beyond a technical focus to embrace a broader health services research perspective on PHR systems are needed to promote further adoption, enhance patient-centered care, and realize the anticipated potential for health system transformation. Health services researchers are in a unique position to address this gap in the evidence base, but doing so will require careful attention to the formulation of research questions and study designs, the prioritization of research areas, and an accounting of the unique factors inherent in PHR research.

In 2003, the Department of Veterans Affairs (VA) introduced My HealtheVet (MHV), a web-based PHR intended to complement traditional services, improve co-managed care, and empower patients and their families to play a more active role in veterans' health. Below we describe the MHV PHR and efforts to evaluate the impact of its use on veterans and the VA healthcare system. Drawing upon the VA experience, we describe salient PHR research questions and potential study design issues. We identify important factors inherent to PHR research that we have thus far identified, and offer lessons learned that can inform research and evaluation efforts surrounding PHR systems across different contexts.

BACKGROUND

My HealtheVet System Overview

My HealtheVet (http://www.myhealth.va.gov) is an integrated PHR that includes health information entered by Veterans, data from VA's unified EHR-S, health education information, health management tools, and links to other resources.[31–36] The system represents collaborative work between multiple offices in VA. The Veterans and Consumers Health Informatics Office (V/CHIO), a division of the Chief Health Informatics Office (CHIO), identifies strategic priorities, coordinates with policy experts, and translates MHV goals into business, functional, and technical requirements, based on veteran and consumer needs and preferences. Other offices perform the technical lifecycle tasks of requirements management, system development, testing, and run-time operations. Hereafter, we refer to this partnership as the MHV Program Office, a functional model that represents the full range of strategic and technical activities. The MHV Program Office is advised by a multidisciplinary Clinical Advisory Board (MHV CAB). From an organizational perspective, this approach yields multiple benefits, including the direct alignment of program goals and resultant PHR design and development strategies with the overarching objectives of VA and its partners.

There are three levels of MHV access, with a progressive increase in functionality (see Table 1). First, portions of MHV can be accessed by anyone with an Internet connection. Second, veterans can create an account by performing an online registration, which provides them with functions not available to the general public. Third, veterans who choose to complete a onetime in-person identity verification at a VA medical center, referred to as "in-person authentication" (IPA), can also view a growing array of additional information extracted from the VA EHR-S. As of July 2009, MHV has been visited over 28 million times, more than 810,000 people have registered (16.3% of veterans currently receiving VA healthcare services), and over 130,000 veterans have completed the IPA process.[37,38] Veteran feedback obtained through the American Customer Satisfaction Index (ACSI) Survey, an industry standard tool for measuring satisfaction and prioritizing improvements,[39] is used to guide system redesign and the addition of new features.

Developing an Evaluation Approach for MHV

The MHV Program Office is working to pursue a robust PHR evaluation program that moves beyond studies focused solely on technological or system concerns to those that reflect a broader health services perspective. Inherent in this shift is a fuller accounting of the social, clinical, and organizational contexts in which PHRs are used. Each of these dimensions is important to identifying optimal PHR features and assessing system impact. Several MHV-related projects initiated by members of the VA research community were not formally designed as part of the Program Office's evaluation effort, but together reflect commitment to an encompassing approach. For example, the MHV Program Office has partnered with the Stroke Quality Enhancement Research Initiative (QUERI) to create age-appropriate, culturally relevant materials for caregivers

Table 1. My HealtheVet Personal Health Record Features

My HealtheVet Personal Health Record feature key: V = all site visitors R = registered users A = authenticated users (IPA)	V	R	A
General information and resources: Access information about Federal and VA benefits and resources, VA-related news and events. Link to additional resources	X	X	X
Research health: Browse and search collections of evidence-based health information including Healthy Living Centers, Condition Centers, and medical databases. Access health screening tools, mental health resources, and articles	X	X	X
My HealtheVet Learning Center: Take online courses to promote mental health	X	X	X
Personal information: Store and maintain contact information including emergency contacts. Manage account profile, preferences, and options		X	X
Get care: Store and maintain information pertaining to caregivers and providers, treatment facilities and locations, and health insurance coverage		X	X
Health information card: Print selected personal and medical information on a pre-formatted wallet card for a convenient reference		X	X
Personal health history: Record important health history information and events		X	X
Family health history: Record family member's health history and events that may affect health		X	X
Military health history: Record important events from military service including assignments related to health history, potential exposures, and treatments		X	X
Personal health summary: Select information to print out as a personal health summary report to share with providers		X	X
Health eLogs: Track and graph common health measures (blood pressure, blood sugar, cholesterol, body temperature, weight, heart rate, pain, pulse oximetry, INR)		X	X
Allergies: Record allergies by date, severity, reaction, diagnosis, and add comments		X	X
Immunizations: Record the immunization, date, method used, and any reactions		X	X
Tests: Record tests by test name, date of test, location where the test was performed, provider's name, results, and add comments		X	X
Medical events: Keep track of illnesses, accidents, or other events by logging the date, treatment prescribed, and any comments regarding the event		X	X
Food and activity journals: Record food intake to monitor diet or control weight, and keep track of exercise routines. Print journal worksheets for easy tracking		X	X
Health calendar: Add events, set reminders, utilize a to-do list		X	X
Medications, over-the-counter drugs, herbals, and supplements: Record the name, starting and ending date, prescription number, and dosage		X	X
Prescription refills: Request refills for VA prescriptions online (authenticated users can view medication names when ordering refills)		X	X
VA prescription history: View a record of all VA prescriptions			X
My complete medications: View and print a complete summary of both VA and self-entered medications to support medication reconciliation			X
Wellness reminders: View customized reminders for preventative care and screens			X
Secure messaging: Exchange secure electronic messages with your healthcare team for non-urgent needs (currently available at 8 sites with further expansion planned)			X

of stroke survivors.[40] Other research studies are currently underway, including studies focused on usability testing,[41] use of MHV as a communication tool for health screening,[42] development of a MHV Healthy Living Center dedicated to spinal cord injury,[43] and the integration of evidence-based tools to assess heart failure care according to published VA guidelines.[44] As we discuss below, close working relationships between PHR system providers and researchers can be instrumental in sustaining such efforts.

We recognized the need for a broader organizing framework in order for VHA to utilize its significant research expertise to optimize the MHV system. The Performance Evaluation Workgroup of the MHV CAB, whose task it is to guide system evaluation efforts, is currently using an extension of the RE-AIM (Reach, Efficacy/Effectiveness, Adoption, Implementation, and Maintenance) framework[45] as a model for assessing the impact of MHV on the veteran population. This framework supports a comprehensive evaluation effort and has resulted in the identification of a number of high priority research areas as shown in Table 2. These range from extending the reach of the program in the veteran population to examining cohorts of users with respect to utility, outcomes, and cost. Ongoing efforts to evaluate the MHV PHR have revealed a number of insights that can inform similar work regarding other PHR systems. In the remainder of this paper, we discuss these lessons and offer examples from our work with MHV, as appropriate.

LESSONS LEARNED FROM A HEALTH SERVICES RESEARCH PERSPECTIVE ON PHRS

Posing Research Questions and Designing Studies

Our experiences underscore the point that many kinds of research questions can be posed in relation to PHRs. Organizations that provide PHRs are generally interested in knowing who is using the system and for what purpose(s), how the design affects use, and whether the content has perceived value. Health services and other researchers, however, are likely to also be interested in additional aspects, including how PHRs can be used to improve patient outcomes, eliminate health disparities, and deliver interventions.

PHR-related studies may entail designs that require engaging PHR users as study participants, or, with their consent, accessing the data that they store in their PHR (e.g., blood pressure readings) as a means of monitoring outcomes. Delineating between research for the purpose of PHR evaluation and studies that utilize the PHR in the course of an intervention is critical. Each approach requires different methodologies and types of collaboration to ensure that goals are clearly identified, and that adequate safeguards are applied to support user needs and expectations. In the case of MHV, the Program Office has thus far emphasized research on users, expecting to address the issue of delivering evidence-based interventions through the system in the future. The scenarios

Table 2. MHV PHR Research Agenda by RE-AIM Domain

RE-AIM domain	MHV program goal	Research priorities
Reach: the number, proportion, and representativeness of individuals who utilize the MHV PHR	Increase enrollment of the veteran population served by the MHV program	• Assess level of awareness in the veteran population and identify effective improvement strategies • Characterize users in comparison with the veteran and VA patient populations • Identify target populations who can most benefit from use
Effectiveness: the impact of MHV PHR utilization on users, outcomes, performance, and organizational systems	Utilize MHV to enhance access to services, improve behavioral and health outcomes, improve quality, increase satisfaction, and enhance system efficiency	• Validate and extend initial analytic findings from user surveys • Examine cohorts of users to evaluate impact on utilization management (access), behavioral and clinical outcomes (quality), cost (value), and satisfaction • Identify how MHV can enhance access to services for rural or special populations • Evaluate impact on workflow, workload, VA performance measures, and organizational processes
Adoption: use of the MHV PHR by veterans and their caregivers, healthcare providers, and healthcare teams	Increase adoption of MHV by veterans, providers, and healthcare teams	• Elicit perceptions of veteran and provider nonadopters • Identify barriers to adoption and develop strategies to address • Examine the current authentication process and identify ways to improve
	Increase levels of engagement and activation with MHV among patients and providers	• Demonstrate clinical utility of MHV components • Compare clinical adoption across settings • Identify communication and process strategies for integration of MHV within clinical practice
Implementation: the efforts and costs involved in implementing the MHV Program	Implement the MHV program nationwide in the most equitable, effective, and efficient manner; enhance program value	• Examine how to prevent further health disparities by exploring issues of access, health literacy, and computer literacy • Evaluate cost/benefit impact of MHV use • Identify optimal implementation strategies
Maintenance: the long-term effects of MHV PHR use and program sustainability	Utilize MHV to enhance and sustain desirable long-term outcomes	• Track long-term impacts on targeted outcomes (access, quality, value, satisfaction) • Identify effective engagement strategies to support sustained MHV use • Explore program enhancements (usability) and expansion (interoperability) • Analyze MHV program sustainability

provided in Table 3 illustrate different ways that a PHR could be employed in intervention studies, either as a vehicle through which to deliver an intervention or as the intervention itself.

A related and particularly challenging issue that warrants mention is that most PHRs are "live systems" with corresponding populations of users. As such, researchers must develop innovative study designs to investigate continually evolving PHR systems and contexts of use. In the time it takes to publish study results, a PHR may have undergone significant changes in terms of content and functionality. Using a PHR to deliver an intervention may also have the effect of making the system even more dynamic and difficult to characterize.

Addressing Issues of Data Ownership and Consent

In contrast to EHR systems, which are predominantly owned and operated by the system provider, PHRs are based on consumer control and management of their own data. In the

Table 3. Potential Scenarios for Research Involving PHRs

Study type	Description
1. Accessing data stored within the PHR	A researcher would like to link information from the PHR with patient medical records. In particular, he is interested in looking at blood pressure rates over time, as entered by patients into the PHR, with patient consent, and linking this information to pharmacy and laboratory data for particular medications with the goal of relating patient outcome (blood pressure control) with medication use
2. Delivering an intervention through the PHR	A researcher is interested in improving patient self-management of diabetes through a healthy diet, regular exercise, and regular monitoring of blood sugar. She wants to provide a series of self-help exercises and educational materials through the PHR
3. Utilizing the PHR as an intervention	Through the use of secure messaging, the PHR could be used to facilitate adherence with requirements for elective surgery (e.g., details for arrival check-in, preoperative instructions such as stopping certain medications)

case of MHV, the overall system is managed by the VA, but the content of the PHR is the property of the veteran. Privacy and security are paramount, and researchers who design studies that include access to PHR data must explicitly obtain the informed consent of consumers. These consent processes must recognize the PHR user as the data owner, protect the integrity of PHR data, and still offer PHR users the opportunity to participate in research intended to enhance the system, improve user experience, or strengthen positive outcomes based on effective use.

Engagement of PHR users in research must also ensure that participation is voluntary and that participants have a clear understanding of the level of data-sharing expected. Organizational policy development must include a review of formal agreements between PHR providers and PHR users regarding the maintenance and protection of data, including the PHR terms and conditions, privacy policy, and system of records. From a technical perspective, the development of standardized processes to support access to data is crucial for future research efforts. For example, analysis of application activity logs could reveal patterns of activity independent of user identity, facilitating key insights about PHR usage. For studies that involve PHR users as consenting participants, user delegation of access to specific PHR data may allow researchers to draw upon patient self-reported data and link it to patient data from medical records and other sources to evaluate outcomes. At VA, this delegation functionality has been successfully piloted with an earlier prototype; however, it is not yet available in the national MHV PHR.

Promoting Effective Working Relationships

Efforts to foster PHR research reveal the importance of multidisciplinary collaboration. Processes and policies must attend to the varying objectives, needs, and requirements of PHR providers and researchers to enable effective collaboration. To the extent possible, approaches to collaboration should leverage existing institutional research policies, structures, and processes rather than recreating or duplicating them, accounting as necessary for nuances specific to PHRs. Development of a research agenda to identify high priority areas of study, similar to that shown in Table 2 for MHV, is one way to address this issue. Additionally, an organizational infrastructure is necessary to develop sustainable research collaborations. Such infrastructure can help guide and support research in ways that align with the vision of system designers and stakeholders while preserving the trust of PHR users.

Discussions with VA researchers have also identified practical tools which can further support research, such as the incorporation of a survey engine within MHV to efficiently host survey research. The development and dissemination of a standardized data dictionary has also been proposed. Such a tool would describe relevant data elements important to PHR research, along with any data-specific constraints. A study currently funded by VA is identifying the data elements and technical infrastructure needed to support MHV research.[46] For each type of PHR data, formal processes must be established in order to enable appropriate access for Institutional Review Board (IRB)-approved studies.

CONCLUSION

Significant further work is needed to understand the use of PHRs as integrated tools that complement traditional care, and to identify the impact of their use on patients, providers, organizations, and healthcare systems. In particular, identifying the effect of PHRs on clinical, behavioral, and financial outcomes will be critical in fostering the cultural transformation and uptake needed to make PHRs an integral part of the fabric of healthcare. These interests are well aligned with health services research. Organizations that offer a PHR or intend to develop one will benefit from elucidating specific PHR research priorities, identifying and addressing research barriers, and finding pragmatic ways to support research efforts.

The lessons that we have thus far learned from our efforts to study MHV are not unique to the system itself or to the VA as an institution. PHRs are new tools intended to support patient-centered healthcare. As such, priority must be given to issues inherent to PHRs, including data ownership, access, privacy, and confidentiality. Organizational policy development must be guided by emerging national privacy policy frameworks, provisions, and laws.[47,48] Technical solutions to foster effective research programs must be driven by organizational policy that ensures adequate protection for users while enabling rigorous investigations. Collaborative approaches that connect PHR system providers with the skills and expertise embodied in research communities are essential to support studies that will optimize PHRs and their use. At the center of this work is the PHR user. As researchers and PHR providers, we must offer clear information about data management policies, privacy and security policies, analysis procedures, and opportunities to participate in research, all while maintaining the integrity of consumer trust. Only in this way can we enable a deeper understanding of the PHR as a contemporary tool and a potentially transformative force in health care.

Conflicts of Interest: *None.*

Corresponding Author: *Kim M. Nazi, PhD(c); Veterans and Consumers Health Informatics Office, Office of Health Information, Veterans Health Administration, US Department of Veterans Affairs, Washington, DC, USA (e-mail: kim.nazi@va.gov).*

REFERENCES

1. Committee on Quality Health Care in America, Institute of Medicine. Crossing the quality chasm: a new health system for the 21st century. Washington, DC: National Academy Press; 2001.
2. **Calabretta N.** Consumer-driven, patient-centered health care in the age of electronic information. J Med Libr Assoc. 2002;90(1):32–7.
3. **Klein-Fedyshin M.** Consumer health informatics – integrating patients, providers, and professionals online. Med Ref Serv Q. 2002;21(3):35–50.
4. **Ball MJ, Lillis J.** E-health: transforming the physician/patient relationship. Int J Med Inform. 2001;61(1):1–10.
5. **Tang PC, Lansky D.** The missing link: bridging the patient-provider health information gap. Health Aff. 2005;24(5):1290–5.
6. **Ball MJ, Gold J.** Banking on health: personal records and information exchange. J Healthc Inf Manag. 2006;20(2):71–83.
7. **Cronin C.** Personal health records: an overview of what is available to the public. 2006. Available at: http://assets.aarp.org/rgcenter/health/2006_11_phr.pdf. Accessed Feb 28, 2009.
8. **Endsley S, Kibbe DC, Linares A, Colorafi K.** An introduction to personal health records. Fam Pract Manag. 2006;13(5):57–62.

9. **Heubusch K.** Piecing together the PHR. J AHIMA. 2007;78(4):28–32.
10. **Sprague L.** Personal health records: the people's choice? NHPF Issue Brief. 2006;820:1–13.
11. **Tang PC, Ash JS, Bates DW, Overhage JM, Sands DZ.** Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption. J Am Med Inform Assoc. 2006;13(2):121–6.
12. **Halamka JD, Mandl KD, Tang PC.** Early experiences with personal health records. J Am Med Inform Assoc. 2008;15(1):1–7.
13. AHIMA e-HIM Personal Health Record Work Group. Defining the personal health record. J AHIMA. 2005;76(6):24–5.
14. **Gearon C.** Perspectives on the future of personal health records. 2007. Available at: http://www.chcf.org/documents/chronicdisease/PHRPerspectives.pdf. Accessed Feb 28, 2009.
15. **Kim MI, Johnson KB.** Personal health records: Evaluation of functionality and utility. J Am Med Inform Assoc. 2002;9(2):171–80.
16. **Kaelber DC, Jha AK, Johnston D, Middleton B, Bates DW.** A research agenda for personal health records (PHRs). J Am Med Inform Assoc. 2008;15(6):729–36.
17. **Sittig DF.** Personal health records on the internet: a snapshot of the pioneers at the end of the 20th century. Int J Med Inform. 2002;65:1–6.
18. **Pagliari C, Detmer D, Singleton P.** Potential of electronic personal health records. Br Med J. 2007;335:330–3.
19. **Eytan T.** Ending secrecy: physician makes case for full disclosure of health records. 2008. Available at: http://www.ihealthbeat.org/Perspectives/2008/Ending-Secrecy-Physician-Makes-Case-for-Full-Disclosure-of-Health-Records.aspx. Accessed Feb 28, 2009.
20. **Greer C, Buck SL.** PHRs and physician practices. J AHIMA. 2007;78(4):71–5.
21. **Hartzband P, Groopman J.** Off the record - avoiding the pitfalls of going electronic. N Engl J Med. 2008;358(16):1656–8.
22. **Steinbrook R.** Personally controlled online health data - the next big thing in medical care? N Engl J Med. 2008;358(16):1653–6.
23. Connecting for Health. Survey finds Americans want electronic personal health information to improve own health care. 2006. Available at: http://www.markle.org/downloadable_assets/research_doc_120706.pdf. Accessed Feb 28, 2009.
24. Harris Interactive. Two in five adults keep personal or family health records and almost everybody thinks this is a good idea. Health Care News. 2004;4(13). Available at: http://www.quickinsights.com/news/newsletters/healthnews/HI_HealthCareNews2004Vol4_Iss13.pdf. Accessed Feb 28, 2009.
25. Harris Interactive. Few patients use of have access to online services for communication with their doctors, but most would like to. 2006. Available at: http://www.harrisinteractive.com/news/allnewsbydate.asp?NewsID = 1096. Accessed Feb 28, 2009.
26. Connecting for Health. Americans overwhelmingly believe electronic personal health records could improve their health. 2008. Available at: http://www.connectingforhealth.org/resources/ResearchBrief-200806.pdf. Accessed Feb 28, 2009.
27. Connecting for Health. The personal health working group: final report. 2003. Available at: http://www.connectingforhealth.org/resources/final_phwg_report1.pdf. Accessed Feb 28, 2009
28. Brennan PF, Downs S, Casper G, Kenron D. Project HealthDesign: stimulating the next generation of personal health records. In: AMIA Annu Symp Proc; 2007:70–4.
29. Connecting for Health. Connecting consumers: a common framework for networked personal health information. 2006. Available at: http://connectingforhealth.org/commonframework/docs/P9_NetworkedPHRs.pdf. Accessed Feb 28, 2009.
30. National Committee on Vital and Health Statistics. Letter to department of Health and Human Services Secretary Michael Leavitt. 2005. Available at: http://www.ncvhs.hhs.gov/050909lt.htm. Accessed Feb 28, 2009.
31. **Schneider J.** My HealtheVet: fighting for health with information. Journal of Consumer Health on the Internet. 2008;12(1):13–21.
32. **Naditz A.** Telemedicine at the VA: VistA, MyHealtheVet, and other VA programs. Telemed J E Health. 2008;14(4):330–2.
33. **Nazi K, Cowell F.** My HealtheVet: VA's new web-based tool helps veterans stay healthy-and involved in their care. PN. 2007;61(9):30–4.
34. **Buxbaum P.** Envisioning troop-centric health care. Government Health IT. 2008 Jun 8. Available at: http://www.govhealthit.com/print/4_18/features/350395-1.html. Accessed October 15, 2008.
35. **Nazi K.** "My HealtheVet Personal Health Record". Testimony to AHIC CE Workgroup, September 12, 2007. Available at: http://www.dhhs.gov/healthit/ahic/materials/09_07/ce/nazi.html. Accessed October 15, 2008.
36. **Nazi K.** "My HealtheVet Personal Health Record, Part II", Testimony to AHIC CE Workgroup, October 16, 2007. Available at: http://www.dhhs.gov/healthit/ahic/materials/10_07/ce/nazi.html. Accessed October 15, 2008.
37. MHV Program Office. My HealtheVet Product Intranet Website, Statistics. 2009.
38. MHV Program Office. My HealtheVet Produet Intranet Website, IPA Statistics. 2009.
39. American Customer Satisfaction Index. ACSI methodology. 2006. Available at: http://www.theacsi.org/index.php?Itemid=41&id=48&option=com_content&task=view. Accessed Feb 23, 2009.
40. MHV Program Office. My HealtheVet Research Summary, September 2008.
41. **Chumbler NR.** Factors influencing implementation of My HealtheVet: final report for HSR&D project SHP 08-192. Washington, DC: Department of Veterans Affairs, Veterans Health Administration; 2008.
42. **McInnes DK, Asch SM, Bokhour BG, et al.** "My HealtheVet for Communication & Social Marketing," Veterans Affairs Quality Enhancement Research Initiative (QUERI) National Meeting, Phoenix, AZ, December 12, 2008.
43. **Weaver FM, Hogan TP, Smith BM, Wallen ES, Goldstein B, Hammond MC.** Using the Web to promote self-management among veterans with spinal cord injuries. In: Congress of Spinal Cord Medicine and Rehabilitation Conference. Orlando, FL; 2008. Abstract available at: http://www.spinaleordcongress.org/pdf/Abstracts%20Final%202008.pdf. Aecessed Feb 28, 2009.
44. **Heidenreich PA.** MHV: health status measure and quality of care assessment: rapid response project proposal 07-275. Washington, DC: Department of Veterans Affairs, Veterans Health Administration; 2008.
45. **Glasgow RE, Linnan LA.** Evaluation of theory-based interventions. In: Glanz K, Rimer BK, Viswanath K, eds. Health behavior and health education: theory, research, and practice. 4th ed. San Francisco: Jossey-Bass; 2008:487–508.
46. **Wagner TH, Weingardt K.** Supporting the evaluation of the My HealtheVet personal health record: rapid response project proposal 08-396. Washington, DC: Department of Veterans Affairs, Veterans Health Administration; 2008.
47. Office of the National Coordinator for Health Information Technology. US Department of Health and Human Services. *Nationwide Privacy and Security Framework for Electronic Exchange of Individually Identifiable Health Information.* 2008. Available at: http://www.hhs.gov/healthit/documents/NationwidePS_Framework.pdf. Accessed Mar 8, 2009.
48. Health Information Technology for Economic and Clinical Health Act (HITECH). Public Law 111-5, Section 13001 and following, 2009.

Addiction
RESEARCH REPORT
doi:10.1111/j.1360-0443.2010.03007.x

US state alcohol sales compared to survey data, 1993–2006

David E. Nelson[1,2], Timothy S. Naimi[2], Robert D. Brewer[2] & James Roeber[3]

Cancer Prevention Fellowship Program, Center for Cancer Training, National Cancer Institute, Bethesda, MD, USA,[1] Alcohol Team, Emerging Investigations and Analytic Methods Branch, Division of Adult and Community Health, National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention (CDC), Atlanta, GA, USA[2] and New Mexico Department of Health, Albuquerque, NM, USA[3]

ABSTRACT

Aims Assess long-term trends of the correlation between alcohol sales data and survey data. **Design** Analyses of state alcohol consumption data from the US Alcohol Epidemiologic Data System based on sales, tax receipts or alcohol shipments. Cross-sectional, state annual estimates of alcohol-related measures for adults from the US Behavioral Risk Factor Surveillance System using telephone surveys. **Setting** United States. **Participants** State alcohol tax authorities, alcohol vendors, alcohol industry (sales data) and randomly selected adults aged ≥ 18 years 1993–2006 (survey data). **Measurements** State-level per capita annual alcohol consumption estimates from sales data. Self-reported alcohol consumption, current drinking, heavy drinking, binge drinking and alcohol-impaired driving from surveys. Correlation coefficients were calculated using linear regression models. **Findings** State survey estimates of consumption accounted for a median of 22% to 32% of state sales data across years. Nevertheless, state consumption estimates from both sources were strongly correlated with annual r-values ranging from 0.55–0.71. State sales data had moderate-to-strong correlations with survey estimates of current drinking, heavy drinking and binge drinking (range of r-values across years: 0.57–0.65; 0.33–0.70 and 0.45–0.61, respectively), but a weaker correlation with alcohol-impaired driving (range of r-values: 0.24–0.56). There were no trends in the magnitude of correlation coefficients. **Conclusions** Although state surveys substantially underestimated alcohol consumption, the consistency of the strength of the association between sales consumption and survey data for most alcohol measures suggest both data sources continue to provide valuable information. These findings support and extend the distribution of consumption model and single distribution theory, suggesting that both sales and survey data are useful for monitoring population changes in alcohol use.

Keywords Alcohol abuse, alcohol drinking, cross-sectional studies, drinking behaviors, health surveys, statistics.

Correspondence to: David E. Nelson, National Cancer Institute; 6120 Executive Boulevard, Suite 150E, MSC 7105, Bethesda, MD 20892-7105, USA.
E-mail: nelsonde@mail.nih.gov
Submitted 5 October 2009; initial review completed 1 December 2009; final version accepted 1 March 2010

INTRODUCTION

Given the enormous health, economic and other costs associated with excessive alcohol use [1,2], monitoring alcohol consumption and other alcohol use behaviors on a population-wide basis is important for surveillance, epidemiological, programmatic and other purposes [1,3–6]. The two most common methods for obtaining population-level measures of alcohol use are per capita consumption estimates based generally on alcohol sales, taxation or product shipment data [1,3] (referred to hereafter as sales data), and self-reported measures based on surveys of individuals [3,4].

Per capita alcohol consumption based on alcohol sales is a widely used measure in alcohol research [1,3], and it is included as a key surveillance measure in the Healthy People 2010 objectives for the United States [6]. A substantial body of evidence from many countries has shown an association between population-based per capita alcohol consumption and adverse health effects of

Disclaimer: The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention or the National Institutes of Health.

alcohol use, including cirrhosis and other alcohol-specific conditions, motor vehicle crashes and suicide [7–17].

Alcohol sales data have several advantages: they are obtained readily and inexpensively; they are measured relatively uniformly across states; they are not affected by recall bias as survey data may be; and they can broadly measure the impact of policies or other interventions across an entire population [3,4,7,16,18,19]. The major disadvantage of sales data is that consumption is averaged across an entire population (including non-drinkers), making it impossible to characterize alcohol consumption patterns within a population or to compare alcohol consumption by factors such as age, sex, race/ethnicity or substate geographic area [4,18,19].

In contrast, survey data based on self-reports can identify drinking patterns readily among population groups and are useful for measuring the potential differential impacts of interventions, such as policy changes or educational programs, on these subgroups [3,4,19]. The major disadvantage is that survey data substantially underestimate alcohol consumption and other alcohol use measures, such as binge drinking and alcohol-impaired driving [3,4,19–21]; the extent of this underestimation may be increasing because of declining survey response rates and non-coverage bias, e.g. people in cellphone-only households are not included in landline-only telephone surveys [22]. For example, people in cellphone-only households are much more likely to be younger adults and to binge drink [22,23].

Somewhat surprisingly, relatively few US studies have directly examined the association between population-based per capita alcohol consumption based on state sales data with self-reported alcohol measures from state surveys [24–26], and none have done so using data collected within the past 10 years. Smith and colleagues [26] reported that survey data underestimated per capita consumption substantially based on sales data among 21 states participating in the 1985 Behavioral Risk Factor Surveillance System (BRFSS), with the median state alcohol consumption estimate averaging only 25% of per capita consumption based on sales data. Nevertheless, state sales data were correlated strongly with survey estimates of annual consumption, heavy drinking, binge drinking and alcohol-impaired driving.

Fitzgerald & Mulford [27] found that from 1985 to 1989, per capita alcohol consumption based on sales data and drinking frequency based on survey data both declined in most locations in Iowa. Rogers & Greenfield [28] pooled data from five US telephone surveys from 1989–94 conducted in 48 US states and the District of Columbia and performed state-specific analyses comparing sales with survey data. They found that, overall, survey data accounted for 35% of sales data (coverage rate), with state estimates ranging from 8% in Wyoming to 60% in North Dakota. They also reported that the correlation coefficient of state coverage rates with state current alcohol use prevalence was 0.33.

To our knowledge, Smart and colleagues [18] have conducted the only study examining the association between sales and survey data over time. Using data from Ontario, Canada from 1977 to 1997, they correlated sales data with survey measures for any drinking in the past year, average number of drinks per week; daily drinking; consuming more than five drinks at a sitting on a weekly basis; and experiencing two or more harmful effects from alcohol use. In contrast to Smith *et al.* [26], daily drinking was the only survey measure that correlated with alcohol sales data ($r = 0.94$) [18].

To assess the validity and utility of both types of population-based measures and to relate alcohol exposure to health outcomes at the population level, it is important to determine if any associations exist currently between state per capita alcohol consumption based on sales and alcohol survey measures, and whether such associations changed over time. To do so, we greatly extended and updated the original Smith *et al.* [26] study and correlated sales-based alcohol consumption data with self-reported survey measures of alcohol consumption, current alcohol use, heavy alcohol use, binge drinking and alcohol-impaired driving from adults aged ≥ 18 years for nearly all 50 US states using data from 1993 through 2006.

METHODS

Per capita alcohol consumption

We obtained state-specific estimates of apparent per capita alcohol consumption from the Alcohol Epidemiologic Data System (AEDS), which is maintained under the auspices of the US National Institute for Alcohol Abuse and Alcoholism (NIAAA). Detailed information about the AEDS is provided elsewhere [29]. Briefly, the AEDS collects and summarizes state and national data on apparent per capita alcohol consumption each year, with public data sets available from 1977 to 2006. Consumption data are based on state sales, state tax receipts or alcohol shipments from alcohol industry sources for beer, wine and spirits. In 2006, for example, the AEDS used sales and/or tax receipts reports from 30 states for beer, 29 states for wine and 26 states for spirits; all other state estimates were based on shipment data.

AEDS uses an estimate of average ethanol content for each type of beverage to convert sales data for the three types of alcohol-containing beverages into a single estimate of total gallons of pure alcohol consumed. To replicate the analyses used by Smith *et al.* [26], denominators

used to estimate per capita consumption from alcohol sales were based on US Census data for people aged ≥ 14 years. Data were not included for the District of Columbia because it is not a state. Analyses of per capita consumption data from the AEDS were limited to the same years for which BRFSS survey data were available.

Survey estimates

Survey data for adult alcohol measures were obtained from publicly available Behavioral Risk Factor Surveillance System data sets for 1993, 1995, 1997, 1999 and 2001–06 from the Centers for Disease Control and Prevention (CDC); details about the BRFSS are available elsewhere [30]. Briefly, the BRFSS obtains state-based estimates on a variety of health risk measures through random digit dial telephone surveys of non-institutionalized people aged ≥ 18 years each year. Overall, the median state sample sizes for the years studied ranged from 2045 to 6080. The state median response rate was 72% in 1993, 69% in 1995, 62% in 1997, 55% in 1999, 51% in 2001, 58% in 2002, 53% in 2003, 53% in 2004, 51% in 2005 and 51% in 2006 [30,31].

We chose 1993 as the starting year for this study for two reasons. First, from 1993 onwards, all 50 states participated in the BRFSS except for Wyoming in 1993 and Hawaii in 2004. Secondly, the wording of BRFSS alcohol-related questions and the definitions of alcohol measures has remained relatively consistent since 1993. The BRFSS contains alcohol questions that covered current use, alcohol consumption (frequency and quantity of alcohol use), binge drinking and alcohol-impaired driving [30]. Only people who reported current alcohol use were asked subsequent alcohol questions. All the alcohol-related questions were asked during each study year with the exception of alcohol-impaired driving, which was included in 1993, 1995, 1997, 1999, 2002, 2004 and 2006.

Current drinking was generally defined as having had at least one alcohol-containing beverage (e.g. beer, wine, wine cooler, liquor) within the past 30 days or past month. Alcohol consumption was based on the average number of days that alcohol was consumed in the past week or month (frequency) and the average number of drinks consumed on those days (quantity). Heavy drinking was defined as consuming an average of two or more drinks per day for men or one or more drink per day for women.

To 2005, binge drinking was defined as consuming more than five drinks on an occasion on one or more of the past 30 days or past month. In 2006, a gender-specific definition of binge drinking was adopted by the BRFSS, based on recommendations from NIAAA [32], and a binge drinking episode was defined as consuming more than four alcohol-containing beverages on an occasion for women (the more than five drinks cut-point remained unchanged for men). Alcohol-impaired driving was defined as having driven more than one time in the past 30 days or past month after having 'perhaps too much to drink'. There were few minor wording changes to questions over time, with the most common change being the use of past 30 days rather than past month as a referent period beginning in 2001.

Analyses

BRFSS state prevalence data were based on the entire population surveyed (i.e. alcohol and non-alcohol users) and were weighted to be representative of each state's adult non-institutionalized population based on the age, sex and race/ethnicity of each state's population using information available from the US Census Bureau. Alcohol consumption and heavy drinking data were not available from California for 1993, 1995 and 1997, because these alcohol-related questions used in that state were different from those used in all other states. For all other years, however, all states used identical alcohol-related questions.

Data from the frequency and quantity questions in the BRFSS were converted to average number of drinks consumed per month, and then multiplied by 12 to estimate the average number of drinks per year. We assumed that one drink of alcohol (regardless of beverage type) contained 0.6 ounces of alcohol [33], multiplied the average annual number of drinks per year by 0.6 ounces, and then divided this number by 128 (the number of ounces in a gallon) to determine per capita annual alcohol consumption in gallons. BRFSS alcohol consumption estimates were divided by AEDS sales data for each state and year to determine the proportion (percentage) of consumption estimated by surveys. We calculated state median, minimum and maximum proportions for each year; similar state data were also obtained on the prevalence of current drinking, binge drinking, heavy drinking and alcohol-impaired driving.

Linear regression models were created to assess the strength of association annually between sales and survey data on (i) average annual per capita alcohol consumption; (ii) current alcohol use prevalence; (iii) binge drinking prevalence; (iv) heavy alcohol use prevalence; and (v) alcohol-impaired driving prevalence. For each model, a β estimate, an *r*-value (correlation coefficient) and an R^2 value were determined, with sales data as the dependent variable and the BRFSS estimate as the independent variable. Correlations of 0.10–0.29 were considered weak, 0.30–0.49 moderate and ≥0.50 strong [34]. Finally, to gain an idea of within-state variability of

Table 1 State sales versus survey data for alcohol consumption estimates, United States, 1993–2006.

Year	*Average annual consumption (gallons): state median and range*		*Survey-based consumption as a percentage of sales-based consumption: state median and range*	*β coefficient (95% CI)*[a]	*r*[b]	R^2
	Sales data	*Survey data*				
1993	2.22 (1.28–4.38)	0.51 (0.25–0.85)	23.3 (13.5–31.5)	0.20 (0.14–0.26)	0.70	0.48
1995	2.19 (1.21–4.12)	0.48 (0.23–0.81)	22.3 (12.6–34.5)	0.16 (0.09–0.23)	0.55	0.29
1997	2.18 (1.25–4.04)	0.51 (0.27–0.83)	23.4 (14.8–30.6)	0.18 (0.12–0.23)	0.67	0.43
1999	2.23 (1.27–4.01)	0.58 (0.28–0.89)	25.0 (15.1–36.7)	0.20 (0.14–0.26)	0.71	0.49
2001	2.23 (1.29–4.00)	0.67 (0.32–1.01)	29.7 (18.0–37.0)	0.21 (0.14–0.27)	0.70	0.48
2002	2.24 (1.31–4.05)	0.71 (0.37–0.97)	31.6 (20.7–40.7)	0.22 (0.14–0.28)	0.66	0.43
2003	2.25 (1.31–4.05)	0.72 (0.33–0.99)	31.6 (16.9–40.7)	0.21 (0.14–0.28)	0.65	0.41
2004	2.27 (1.28–4.07)	0.65 (0.33–0.88)	27.7 (19.0–38.5)	0.19 (0.13–0.25)	0.67	0.44
2005	2.30 (1.26–4.11)	0.64 (0.35–1.17)	28.7 (18.0–38.5)	0.22 (0.15–0.29)	0.68	0.46
2006	2.34 (1.30–4.21)	0.64 (0.30–0.87)	26.4 (17.3–34.8)	0.17 (0.12–0.23)	0.68	0.45

[a]Based on linear regression analyses; $P < 0.001$ for all years. [b]Correlation coefficient. CI: confidence interval.

Table 2 Relationship between state alcohol sales data and survey data for current alcohol use, binge drinking, and heavy drinking, 1993–2006, based on linear regression models.

Year	*Current drinking*			*Binge drinking*			*Heavy drinking*		
	β coefficient (95% CI)[a]	*r*	R^2	*β Coefficient (95% CI)*[a]	*r*	R^2	*β Coefficient (95% CI)*[b]	*r*	R^2
1993	0.13 (0.08–0.18)	0.59	0.33	0.04 (0.02–0.06)	0.56	0.30	0.02 (0.01–0.02)	0.66	0.43
1995	0.14 (0.08–0.19)	0.57	0.31	0.04 (0.02–0.06)	0.47	0.21	0.01 (0.00–0.03)	0.33	0.09
1997	0.14 (0.09–0.20)	0.62	0.37	0.04 (0.02–0.06)	0.49	0.22	0.01 (0.00–0.02)	0.40	0.14
1999	0.13 (0.08–0.18)	0.59	0.34	0.05 (0.03–0.07)	0.61	0.36	0.02 (0.01–0.03)	0.63	0.39
2001	0.14 (0.09–0.19)	0.63	0.38	0.04 (0.02–0.06)	0.51	0.25	0.02 (0.01–0.02)	0.66	0.42
2002	0.14 (0.09–0.19)	0.60	0.35	0.04 (0.02–0.06)	0.51	0.25	0.02 (0.01–0.03)	0.64	0.39
2003	0.14 (0.09–0.19)	0.60	0.35	0.04 (0.02–0.06)	0.56	0.30	0.02 (0.02–0.03)	0.70	0.47
2004	0.13 (0.08–0.18)	0.62	0.37	0.03 (0.02–0.05)	0.54	0.27	0.02 (0.01–0.02)	0.64	0.40
2005	0.13 (0.09–0.18)	0.65	0.41	0.03 (0.02–0.05)	0.56	0.30	0.02 (0.01–0.02)	0.66	0.43
2006	0.12 (0.07–0.16)	0.61	0.36	0.03 (0.01–0.05)	0.45	0.19	0.01 (0.01–0.02)	0.59	0.33

[a]$P < 0.001$ for all years. [b]$P < 0.001$ for all years except 1995 ($P = 0.019$) and 1997 $P = 0.005$). CI: confidence interval.

alcohol consumption estimates from sales and survey data, we calculated quintile rankings in 2006 for alcohol consumption based on both sources.

RESULTS

From 1993 to 2006, the relationship between estimates of per capita alcohol consumption from sales and survey data across states was generally consistent (Table 1), with median state estimates of alcohol consumption from surveys accounting for 22.3–31.6% of estimates based on sales data. However, there was a strong correlation between the two sources of state estimates over time, with correlation coefficients (*r*-values) ranging from 0.55 to 0.71, and β values ranging from 0.16 to 0.22.

The median state prevalence of current, binge and heavy drinking varied little over the study period (range: 52.1–59.3%; 14.1–16.5%; and 3.4–5.8%, respectively; data not shown in tables). We found consistently strong and similar magnitudes of β values and correlation coefficients between state sales-based consumption and survey estimates of current drinking, binge drinking and heavy drinking prevalence (Table 2). Correlation coefficients ranged from 0.57 to 0.65 for current drinking and from 0.45 to 0.61 for binge drinking, with slight yearly variations. There was somewhat greater variability in the correlation between sales data with heavy drinking prevalence (range: 0.33–0.70). The median state prevalence of alcohol-impaired driving also remained relatively unchanged over time (range: 1.9–2.5%; data not shown in tables). Although sales data were correlated with alcohol-impaired driving in all years (Table 3), coefficients were generally lower and had more yearly variation (range: 0.24–0.56).

State estimates and quintile rankings from sales and survey data on alcohol consumption for 2006 are shown

Table 3 Relationship between state alcohol sales data and survey data for alcohol-impaired driving, 1993–2006 based on linear regression models.

Year	*β Coefficient (95% CI)*	*P-value*	*r*	R^2
1993	0.01 (0.01–0.02)	<0.001	0.56	0.28
1995	0.01 (0.00–0.01)	0.079	0.25	0.04
1997	0.01 (0.00–0.01)	0.023	0.32	0.08
1999	0.01 (0.01–0.02)	<0.001	0.55	0.29
2002	0.01 (0.00–0.01)	0.086	0.24	0.04
2004	0.01 (0.00–0.01)	0.002	0.44	0.18
2006	0.01 (0.00–0.01)	0.006	0.39	0.15

CI: confidence interval.

in Table 4. Consumption estimates from surveys as a percentage of sales data ranged from 17.3% in Kentucky to 34.8% in Ohio. States where alcohol consumption estimates from surveys averaged less than 22% from that based on sales were Nevada, Louisiana, Mississippi, New Hampshire and Kentucky. For most states, quintile rankings of alcohol consumption from sales and survey data were similar, but there were a few outliers. For example, Texas was in the 4th quintile based on sales but in the 1st quintile based on survey data; Iowa and Ohio were also in the 4th quintile based on sales but in the 2nd quintile from survey data. By contrast, states such as Louisiana, Idaho and South Dakota were in the 2nd quintile based on sales but in the 4th quintile based on survey data.

DISCUSSION

To our knowledge, this is the first study to examine the association between per capita alcohol consumption estimates from sales data with population-based survey estimates from all 50 states over time. We found that self-reported alcohol consumption from estimates from state surveys accounted for only 22–32% of per capita consumption based on alcohol sales, which was similar to the estimate of 25% found by Smith *et al*. using BRFSS data from 21 states in 1985 [26]. These findings confirm prior research that self-reports substantially underestimate alcohol consumption. However, our estimates were lower than 40–60% as reported by Rehm in a 1998 literature review [19], and 52% as reported by Kerr & Greenfield [35] based on a national comparison of sales data with the 2000 US National Alcohol Survey. They are also lower than the estimate of 35% reported by Rogers & Greenfield based on pooled data from 1989–94 [28]. The use of alcohol-focused surveys, and the inclusion of graduated frequency and other types of questions related to alcohol use, may account for the higher consumption estimates found in these other studies [36].

What was most notable about our results was the striking consistency of the correlations and regression

Table 4 State-specific sales and survey estimates of alcohol consumption, 2006.[a]

State	*Per capita sales (gallons)*	*Survey per capita consumption (gallons)*	*Survey-based consumption as a percentage of sales-based consumption (%)*
New Hampshire	4.21 (1)	0.74 (1)	17.5
Nevada	3.70 (1)	0.81 (1)	21.8
Delaware	3.30 (1)	0.82 (1)	24.8
Wisconsin	2.96 (1)	0.87 (1)	29.2
Montana	2.74 (1)	0.68 (2)	24.8
Wyoming	2.74 (1)	0.71 (2)	25.7
North Dakota	2.74 (1)	0.67 (3)	24.6
Alaska	2.73 (1)	0.71(2)	26.1
Florida	2.73 (1)	0.73 (1)	26.6
Colorado	2.69 (1)	0.64 (3)	23.8
Louisiana	2.66 (2)	0.57 (4)	21.6
Vermont	2.62 (2)	0.85 (1)	32.6
Hawaii	2.56 (2)	0.79 (1)	30.9
Massachusetts	2.55 (2)	0.74 (1)	28.9
Oregon	2.53 (2)	0.68 (2)	27.1
Idaho	2.53 (2)	0.60 (4)	23.6
Rhode Island	2.52 (2)	0.76 (1)	30.2
South Dakota	2.48 (2)	0.55 (4)	22.2
Maine	2.47 (2)	0.71 (2)	28.9
Arizona	2.47 (2)	0.68 (2)	27.4
South Carolina	2.46 (3)	0.55 (4)	22.5
Minnesota	2.40 (3)	0.60 (3)	25.1
New Mexico	2.39 (3)	0.58 (4)	24.3
Missouri	2.38 (3)	0.62 (3)	26.0
Nebraska	2.35 (3)	0.56 (4)	23.8
Illinois	2.33 (3)	0.66 (3)	28.4
New Jersey	2.32 (3)	0.59 (4)	25.3
Connecticut	2.32 (3)	0.67 (3)	29.1
California	2.29 (3)	0.71 (2)	31.0
Mississippi	2.25 (3)	0.42 (5)	18.9
Washington	2.24 (4)	0.67 (3)	29.9
Texas	2.24 (4)	0.76 (1)	34.0
Maryland	2.19 (4)	0.60 (4)	27.2
Iowa	2.19 (4)	0.71 (2)	32.4
Michigan	2.17 (4)	0.68 (2)	31.1
Pennsylvania	2.12 (4)	0.61 (3)	29.0
Virginia	2.10 (4)	0.61 (3)	29.1
Georgia	2.06 (4)	0.50 (5)	24.4
Indiana	2.00 (4)	0.56 (4)	28.0
Ohio	2.00 (4)	0.70 (2)	34.8
New York	1.99 (5)	0.64 (3)	31.9
Alabama	1.97 (5)	0.45 (5)	22.6
North Carolina	1.97 (5)	0.45 (5)	22.7
Kansas	1.94 (5)	0.52 (5)	26.7
Tennessee	1.89 (5)	0.47 (5)	24.7
Kentucky	1.83 (5)	0.32 (5)	17.3
Arkansas	1.82 (5)	0.55 (4)	30.0
West Virginia	1.74 (5)	0.41 (5)	23.4
Oklahoma	1.51 (5)	0.49 (5)	32.4
Utah	1.30 (5)	0.30 (5)	23.1
Median	2.34	0.64	26.4
Range	1.30–4.21	0.30–0.87	17.3–33.8

[a]Numbers in parentheses represent the quintile (from highest to lowest) of each state for alcohol consumption.

coefficients over time between estimates of state per capita consumption from sales data with survey data for current, binge and heavy drinking. Our study's findings were similar to those found in the Smith study for these measures (correlation coefficients of 0.81, 0.74 and 0.59, respectively) [26], although the magnitudes were slightly lower. State alcohol consumption sales data were also correlated with survey estimates of alcohol-impaired driving over time, although the relationship was less strong and consistent; by comparison, Smith *et al.* [26] found a correlation coefficient of 0.51. The consistency of our findings over time suggests that the declining survey response rate and increasing percentage of cellphone households had little effect on the observed correlations, at least to 2006 [37].

Our findings contrast with those of Smart and colleagues based on 1977–97 data from Ontario, who found little correlation between alcohol consumption based on sales data with most survey-based alcohol measures [18]. Reasons for these differences are unclear, but may be the result of different years of study; inclusion of different alcohol survey measures; differences in alcohol consumption patterns; or Smart *et al.*'s analyses of a more limited number of data points from a single province.

The quintile rankings of states for per capita alcohol consumption based on sales were generally similar to those based on survey data, although there were larger differences for a handful of selected states. Reasons for the within-state differences in sales versus survey data for alcohol consumption cannot be gleaned from these data. We speculate that they may stem from cross-border sales (i.e. where large differences in tax policies exist among neighboring states, such as New Hampshire), from non-resident purchases of alcohol-containing beverages in states with high levels of tourism or gaming (e.g. Louisiana, Mississippi, Nevada) or from relatively higher or lower consumption of non-commercially produced alcohol in some states.

The strong relationship we found between per capita consumption estimates based on sales data with specific patterns of alcohol use based on survey data (e.g. binge drinking), as well as the relationship between alcohol sales data with various alcohol-related outcomes observed in prior research [7–17], provide some support for the distribution of consumption model, the single distribution theory and similar frameworks [7,18,38–40]. These frameworks, despite their limitations, posit that alcohol consumption in populations occurs in a relatively fixed distribution such that per capita consumption based on sales data provide a useful proxy for excessive alcohol use and its related harms [18]. For example, in the United States the top 5% of alcohol users account for about 40%, and the top 20% of users for about 90%, of all sales of alcohol-containing beverages [41]. Empirical evidence for these theoretical frameworks is supported further because half the alcohol consumed in the United States is consumed during days where drinkers consume five or more drinks [42].

This study has limitations. For sales data, alcohol sold may not represent actual consumption because people may store beverages and alcohol wastage is not taken into account [3]. Assumptions were made about the average alcohol content for each type of beverage [29], but the volume and percentage of alcohol varies by type of beverage and has changed over time [36]. Alcohol brought into states from other places or made by individuals themselves (e.g. moonshine) is not included. Consumption by tourists or by people on military bases may affect estimates. There were differences in the age groups for denominators used in the two data sources, with sales data based on the population aged $\geq$ 14 years and BRFSS data based on people aged $\geq$ 18 years.

There were additional limitations associated with survey data [4,19], and the BRFSS in particular [30,43]. Under-reporting because of social desirability [44] is a well-known phenomenon. Usual quantity and frequency of alcohol use within the past week or 30 days (month) was projected over a 1-year period. Younger people, particularly men, are individuals who are at greater risk for heavier alcohol use but they are also less likely to have landline telephones or respond to surveys [23]. People on military bases are not part of the BRFSS sample. Estimates of alcohol consumption are based on average alcohol content in beverages. Similar to other telephone surveys [22], there have been declines in response rates to the BRFSS, although the impact of this decline on alcohol measures is not clear. There were slight variations in BRFSS survey questions over time, but this appears to have had little impact. Survey-based consumption estimates would probably have been higher if beverage-specific, graduated frequency or context-specific questions were used [4,19–21,45–47].

We did not use indexing to adjust average alcohol consumption and account for the number of drinks consumed during binge episodes [48,49]. In a study using 2003 BRFSS data on the quantity and frequency of alcohol consumption among US adults, Stahre and colleagues [48] found that indexing average daily alcohol consumption to account for binge drinking episodes increased per capita daily alcohol consumption by 14% and heavy drinking prevalence by 42%. Indexing BRFSS data for binge drinking using the Stahre *et al.* approach from 2003 to 2006 would have increased the median estimates of per capita alcohol consumption to 31–36% of sales data.

Our findings have important implications for public health surveillance on excessive alcohol consumption and related harms, and for the evaluation of population-

based programs to prevent excessive drinking. First, they affirm that both per capita sales and survey data are useful for assessing trends in alcohol consumption and that each provides a somewhat different perspective on alcohol consumption at both the national and state levels. Secondly, despite clear differences in sales- and survey-based estimates of per capita alcohol consumption, these two measures correlate quite well at the state level and can provide a useful basis for planning and evaluating prevention programs in states.

Finally, our findings show that alcohol consumption is even more substantially under-reported than reported elsewhere, thus current BRFSS estimates of key patterns of alcohol consumption such as binge drinking are probably only capturing a third or less of the actual amount of alcohol consumed by US adults. This emphasizes the need to continue monitoring both per capita sales and survey-based measures of alcohol consumption [1,50].

Declarations of interest

None.

Acknowledgements

We thank Paul Mowery and David Wheeler for their assistance.

References

1. Babor T. F., Caetano R., Casswell S., Edwards G., Giesbrecht N., Graham K. *et al. Alcohol: No Ordinary Commodity – Research and Public Policy*. New York: Oxford University Press; 2003.
2. Centers for Disease Control and Prevention. Alcohol-attributable deaths and years of potential life lost—United States, 2001. *Morb Mortal Wkly Rep* 2004; **53**: 866–70.
3. Greenfield T. K., Kerr W. C. Tracking alcohol consumption over time. *Alcohol Res Health* 2003; **27**: 30–8.
4. Dawson D. A. Methodological issues in measuring alcohol use. *Alcohol Res Health* 2003; **27**: 18–29.
5. US Department of Health and Human Services. *Tenth Special Report to the US Congress on Alcohol and Health*. Bethesda, MD: National Institute on Alcohol Abuse and Alcoholism, National Institutes of Health; 2000.
6. US Department of Health and Human Services. *Healthy People 2010. Understanding and Improving Health*, 2nd edn. Washington, DC: US Government Printing Office; 2000.
7. Rush B., Steinberg M., Brook R. The relationships among alcohol availability, alcohol consumption, and alcohol-related damage in the Province of Ontario and the State of Michigan 1955–1982. *Adv Alcohol Subst Abuse* 1986; **5**: 33–45.
8. Xie X., Mann R. E., Smart R. G. The direct and indirect relationships between alcohol prevention measures and alcoholic liver cirrhosis mortality. *J Stud Alcohol* 2000; **61**: 499–506.
9. Ramstedt M. Per capita alcohol consumption and liver cirrhosis mortality in 14 European countries. *Addiction* 2001; **96**: S19–S33.
10. Ramstedt M. Alcohol and suicide in 14 European countries. *Addiction* 2001; **96**: S59–S75.
11. Ramstedt M. Alcohol consumption and liver cirrhosis mortality with and without mention of alcohol—the case of Canada. *Addiction* 2003; **98**: 1267–76.
12. Ramstedt M. Alcohol consumption and alcohol-related mortality in Canada, 1950–2000. *Can J Public Health* 2004; **95**: 121–6.
13. Ramstedt M. Alcohol and pancreatitis mortality at the population level: experiences from 14 western countries. *Addiction* 2004; **99**: 1255–61.
14. Norstrom T., Ramstedt M. Mortality and population drinking: a review of the literature. *Drug Alcohol Rev* 2005; **24**: 537–47.
15. Ramstedt M. Alcohol and suicide at the population level—the Canadian experience. *Drug Alcohol Rev* 2005; **24**: 203–8.
16. Andreasson S., Holder H. D., Norstrom T., Osterberg E., Rossow I. Estimates of harm associated with changes in Swedish alcohol policy: results from past and present estimates. *Addiction* 2006; **101**: 1096–105.
17. Ramstedt M. Alcohol and fatal accidents in the United States—a time series analysis for 1950–2002. *Accid Anal Prev* 2008; **40**: 1273–81.
18. Smart R. G., Suurvali H. M., Mann R. E. Do changes in per capita consumption mirror changes in drinking patterns? *J Stud Alcohol* 2000; **61**: 6225.
19. Rehm J. Measuring quantity, frequency, and volume of drinking. *Alcohol Clin Exp Res* 1998; **22**: 4S–14S.
20. Dawson D. A. Measuring alcohol consumption: limitations and prospects for improvement. *Addiction* 1998; **93**: 965–8.
21. Dawson D. Volume of ethanol consumption: effects of different approaches to measurement. *J Stud Alcohol* 1998; **59**: 191–7.
22. Kempf A. M., Remington P. L. New challenges for telephone survey research in the twenty-first century. *Annu Rev Public Health* 2007; **28**: 113–26.
23. Link M. W., Battaglia M. P., Frankel M. R., Osborn L., Mokdad A. H. Reaching the US cell phone generation: comparison of cell phone survey results with an ongoing landline telephone survey. *Public Opin Q* 2007; **71**: 814–39.
24. Midanik L. The validity of self-reported alcohol consumption and alcohol problems: a literature review. *Br J Addict* 1982; **77**: 357–82.
25. Midanik L. Validity of self-reported alcohol use: a literature review and assessment. *Br J Addict* 1988; **83**: 1019–29.
26. Smith P. F., Remington P. L., Williamson D. F., Anda R. F. A comparison of alcohol sales data with survey data on self-reported alcohol use in 21 states. *Am J Public Health* 1990; **80**: 309–12.
27. Fitzgerald J. L., Mulford H. A. Alcohol availability, drinking contexts and drinking problems: the Iowa experience. *J Stud Alcohol* 1993; **54**: 320–5.
28. Rogers J. D., Greenfield T. K. Are estimates of the concentration of alcohol consumption affected by undercoverage? Evidence from five pooled US surveys. *Contemp Drug Probl* 2000; **27**: 367–81.
29. Lakins N. E., LaVallee R. A., Williams G. D., Yi H. *Apparent Per Capita Alcohol Consumption: National, State, and Regional Trends, 1977–2006*. Surveillance Report no. 85. Bethesda, MD: National Institutes of Health, National Institute on Alcohol Abuse and Alcoholism, Division of Epidemiology and Prevention Research, Alcohol Epidemiologic Data System; 2008.

30. Centers for Disease Control and Prevention [homepage on the Internet] (CDC 2009a). *Behavioral Risk Factor Surveillance System*. Atlanta, GA: CDC; 2009. Available at: http://www.cdc.gov/brfss/ (accessed 9 April 2009).
31. Fahimi M., Link M., Schwartz D. A., Levy P., Mokdad A. Tracking chronic disease and risk behavior prevalence as survey participation declines: statistics from the Behavioral Risk Factor Surveillance System and other national surveys. *Prev Chronic Dis* 2008; **5**. Online publication. Available at: http://www.cdc.gov/pcd/issues/2008/jul/07-0097.htm (accessed 5 January 2010).
32. National Institute of Alcohol Abuse and Alcoholism (NIAAA). NIAAA council approves definition of binge drinking. NIAAA Newsletter 2004; no. 3, p. 3.
33. Centers for Disease Control and Prevention [homepage on the Internet]. *Alcohol & Public Health*. Atlanta, GA: CDC; 2009. Available at: http://www.cdc.gov/alcohol/ (accessed 9 April 2009).
34. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd edn. Hillsdale, NJ: Lawrence Erlbaum; 1988.
35. Kerr W. C., Greenfield T. K. 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**: 1714–22.
36. Greenfield T. K., Kerr W. C. Alcohol measurement methodology in epidemiology: recent advances and opportunities. *Addiction* 2008; **103**: 1082–99.
37. Blumberg S. J., Luke J. V. Reevaluating the need for concern regarding noncoverage bias in landline surveys. *Am J Public Health* 2009; **99**: 1806–10.
38. Cook P. J. *Paying the Tab: The Costs and Benefits of Alcohol Control*. Princeton, NJ: Princeton University Press; 2007.
39. Parker D. A., Harman M. S. The distribution of consumption model of prevention and alcohol problems: a critical assessment. *J Stud Alcohol* 1978; **39**: 377–99.
40. Bruun K., Edwards G., Lumio M., Makela K., Pan L., Popham R. E. *et al. Alcohol Control Policies in Public Health Perspective*, vol. 25. Helsinki, Finland: Finnish Foundation for Alcohol Studies; 1975.
41. Greenfield T. K., Rogers J. D. Who drinks most of the alcohol in the US? *J Stud Alcohol* 1999; **60**: 78–89.
42. Office of Juvenile Justice and Delinquency Prevention. *Drinking in America: Myths, Realities, and Prevention Policy*. Washington, DC: US Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention; 2005.
43. Nelson D. E., Holtzman D., Bolen J., Stanwyck C. A., Mack K. A. Review of the reliability and validity of measures used on the core and rotating core of the Behavioral Risk Factor Surveillance System. *Soc Prev Med* 2001; **46**: S3–42.
44. King M. F., Bruner G. C. Social desirability bias: a neglected aspect of validity testing. *Psychol Mar* 2000; **17**: 79–103.
45. Casswell S., Huckle T., Pledger M. Survey data need not underestimate alcohol consumption. *Alcohol Clin Exp Res* 2002; **26**: 1561–7.
46. Russell M., Welte J. W., Barnes G. M. Quantity–frequency measures of alcohol consumption: beverage-specific vs global questions. *Br J Addict* 1991; **86**: 409–17.
47. Lemmens P., Tane E., Knibbe R. Measuring quantity and frequency of drinking in a general population survey: a comparison of five indices. *J Stud Alcohol* 1992; **53**: 476–86.
48. Stahre M., Naimi T., Brewer R., Holt J. Measuring average alcohol consumption: the impact of including binge drinks in quantity–frequency calculations. *Addiction* 2006; **101**: 1711–8.
49. Goransson M., Hanson B. S. How much can data on days with heavy drinking decrease the underestimation of true alcohol consumption? *J Stud Alcohol* 1994; **55**: 695–700.
50. The Guide to Community Preventive Services. *Preventing Excessive Alcohol Use*. Atlanta, GA: Task Force Community Preventive Services; 2009. Available at: http://www.thecommunityguide.org/alcohol/index.html (accessed 9 August 2009).

Journal of Gerontology: MEDICAL SCIENCES
Cite journal as: J Gerontol A Biol Sci Med Sci
doi:10.1093/gerona/glq028

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

Anne B. Newman[1,*] Stefan Walter[2,3,*,†] Kathryn L. Lunetta[4,5,*] Melissa E. Garcia[6,*] P. Eline Slagboom,[7] Kaare Christensen,[8] Alice M. Arnold,[9] Thor Aspelund,[10] Yurii S. Aulchenko,[2,†] Emelia J. Benjamin[4,11,12,13] Lene Christiansen,[8] Ralph B. D'Agostino Sr[4,14] Annette L. Fitzpatrick,[15] Nora Franceschini,[16] Nicole L. Glazer,[17,18] Vilmundur Gudnason[10,19] Albert Hofman,[2,†] Robert Kaplan,[20] David Karasik,[21,22] Margaret Kelly-Hayes[4,23] Douglas P. Kiel,[21,22] Lenore J. Launer,[6] Kristin D. Marciante,[17,18] Joseph M. Massaro[4,14] Iva Miljkovic,[1] Michael A. Nalls,[24] Dena Hernandez,[24] Bruce M. Psaty[15,17,18,25] Fernando Rivadeneira[2,26,†] Jerome Rotter,[27] Sudha Seshadri,[23] Albert V. Smith,[18] Kent D. Taylor,[27] Henning Tiemeier[2,28,†] Hae-Won Uh,[29] André G. Uitterlinden[2,26,†] James W. Vaupel,[30] Jeremy Walston,[31] Rudi G. J. Westendorp,[32] Tamara B. Harris[6,*] Thomas Lumley[9,*] Cornelia M. van Duijn[2,*,†] and Joanne M. Murabito[4,33,*]

[1]Cardiovascular Health Study, Department of Epidemiology, University of Pittsburgh, Pennsylvania.
[2]Rotterdam Study, Department of Epidemiology and Biostatistics and [3]Department of Public Health, Erasmus Medical Center, Rotterdam, The Netherlands.
[4]Framingham Heart Study, The National Heart Lung and Blood Institute, Massachusetts.
[5]Framingham Heart Study, Department of Biostatistics, Boston University School of Public Health, Massachusetts.
[6]Age, Gene/Environment Susceptibility-Reykjavik Study (AGES-Reykjavik), Laboratory of Epidemiology, Biometry and Demography, Intramural Research Program, National Institute on Aging, Bethesda, Maryland.
[7]Leiden Longevity Study, Netherlands Consortium for Healthy Ageing: Section of Molecular Epidemiology, Leiden University Medical Center, The Netherlands.
[8]Danish 1905 cohort, The Danish Aging Research Center, Institute of Public Health, University of Southern Denmark, Odense, Denmark.
[9]Cardiovascular Health Study, Department of Biostatistics, University of Washington, Seattle.
[10]Age, Gene/Environment Susceptibility-Reykjavik Study (AGES-Reykjavik), University of Iceland School of Medicine, Reykjavik, Iceland.
[11]Framingham Heart Study, Department of Cardiology and [12]Department of Preventive Medicine, Boston University School of Medicine, Massachusetts.
[13]Framingham Heart Study, Department of Epidemiology, Boston University School of Public Health, Massachusetts.
[14]Framingham Heart Study, Statistics and Consulting Unit, Department of Mathematics, Boston University, Massachusetts.
[15]Cardiovascular Health Study, Department of Epidemiology, University of Washington, Seattle.
[16]Atherosclerosis Risk in Communities Study, Department of Epidemiology, University of North Carolina, Chapel Hill.
[17]Cardiovascular Health Study, Cardiovascular Health Research Unit, University of Washington, Seattle.
[18]Department of Medicine, University of Washington, Seattle.
[19]Age, Gene/Environment Susceptibility-Reykjavik Study (AGES-Reykjavik), Icelandic Heart Association, Iceland.
[20]Cardiovascular Health Study, Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, New York.
[21]Framingham Heart Study, Hebrew Senior Life Institute for Aging Research, Boston, Massachusetts.
[22]Harvard Medical School, Boston, Massachusetts.
[23]Framingham Heart Study, Department of Neurology, Boston University School of Medicine, Massachusetts.
[24]Age, Gene/Environment Susceptibility-Reykjavik Study (AGES-Reykjavik), Laboratory of Neurogenetics, Intramural Research Program, National Institute on Aging, Bethesda, Maryland.
[25]Cardiovascular Health Study, Group Health Research Institute, Group Health Cooperative, Seattle, Washington.
[26]Rotterdam Study, Department of Internal Medicine, Erasmus Medical Center, Rotterdam, The Netherlands.
[27]Cardiovascular Health Study, Medical Genetics Institute, Cedars-Sinai Medical Center, Los Angeles, California.
[28]Rotterdam Study, Department of Child and Adolescent Psychiatry, Erasmus Medical Center and Sophia Children's Hospital, Rotterdam, The Netherlands.
[29]Leiden Longevity Study, Netherlands Consortium for Healthy Ageing: Section of Medical Statistics, Leiden University Medical Center, The Netherlands.
[30]Danish 1905 cohort, Population, Policy and Aging Research Center, Duke University, Durham, North Carolina.
[31]Cardiovascular Health Study, Department of Medicine, Johns Hopkins University, Baltimore, Maryland.
[32]Leiden Longevity Study, Department of Gerontology, Leiden University Medical Center, The Netherlands.
[33]Framingham Heart Study, Section of General Internal Medicine, Boston University School of Medicine, Massachusetts.

*These authors contributed equally to this work.

†Member of the Netherlands Genomics Initiative–sponsored Netherlands Consortium for Health Aging.

Address correspondence to Anne B. Newman, MD, MPH, Department of Epidemiology, Graduate School of Public Health, 130 North Bellefield Avenue, Suite 500, University of Pittsburgh, Pittsburgh, PA 15213. Email: newmana@edc.pitt.edu

Background. Genome-wide association studies (GWAS) may yield insights into longevity.

Methods. We performed a meta-analysis of GWAS in Caucasians from four prospective cohort studies: the Age, Gene/Environment Susceptibility-Reykjavik Study, the Cardiovascular Health Study, the Framingham Heart Study, and the Rotterdam Study participating in the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium. Longevity was defined as survival to age 90 years or older (n = 1,836); the comparison group comprised cohort members who died between the ages of 55 and 80 years (n = 1,955). In a second discovery stage, additional genotyping was conducted in the Leiden Longevity Study cohort and the Danish 1905 cohort.

Results. There were 273 single-nucleotide polymorphism (SNP) associations with $p < .0001$, but none reached the prespecified significance level of 5×10^{-8}. Of the most significant SNPs, 24 were independent signals, and 16 of these SNPs were successfully genotyped in the second discovery stage, with one association for rs9664222, reaching 6.77×10^{-7} for the combined meta-analysis of CHARGE and the stage 2 cohorts. The SNP lies in a region near *MINPP1* (chromosome 10), a well-conserved gene involved in regulation of cellular proliferation. The minor allele was associated with lower odds of survival past age 90 (odds ratio = 0.82). Associations of interest in a homologue of the longevity assurance gene (*LASS3*) and *PAPPA2* were not strengthened in the second stage.

Conclusion. Survival studies of larger size or more extreme or specific phenotypes may support or refine these initial findings.

Key Words: Longevity—Genome-wide association study—Meta-analysis.

Received September 22, 2009; Accepted February 15, 2010

Decision Editor: Luigi Ferrucci, MD, PhD

INCREASES in longevity of the general population worldwide are an unprecedented phenomenon with significant health and social impact. Although environmental factors have led to an increase in life span, there is ample evidence that genetic factors are involved in extreme longevity both in humans (1–7) and in other organisms (8). The protective genetic factors that lead to longevity are likely to involve fundamental processes of aging that may be different from those associated with early mortality or premature onset of age-related diseases in younger individuals. The mechanisms of aging in humans are far from understood, but available evidence suggests that several pathways—inflammation, oxidative stress and stress responses, cellular senescence, DNA damage and repair, and the growth hormone or insulin-like growth factor and insulin (GH, IGF, INS) axis—may play key roles (9–12). Model organisms suggest that inhibiting the GH, IGF, or INS axis, which is involved in regulating cell proliferation, cell death, wound repair, and metabolism, may promote longevity by reducing oxidative stress and slowing the rate of cell replication and the accumulation of somatic-cell DNA mutations (13). There is also evidence for other important pathways such as the heat-shock proteins and heat-shock factors that are highly conserved across species and play a role in prolongevity transcription pathways. Clinical and epidemiological investigations, including candidate gene studies, have suggested that inflammation pathways may affect life span and risk of age-related conditions such as cardiovascular disease (CVD) and its risk factors (14–19). A combination of multiple genetic variants may be required for an individual to achieve exceptional longevity, which may account in part for its rarity.

Two previous studies have used whole-genome screening to identify genetic variants associated with longevity (20,21). In a linkage analysis, the earliest report (20) identified a locus on chromosome 4 that has not been replicated. A recent report from the Framingham Heart Study (FHS) (22) identified modest associations between longevity (or age at death) and single-nucleotide polymorphisms (SNPs) in or near important candidate genes, including *FOXO1A*, *GAPDH*, *KL*, *LEPR*, *PON1*, *PSEN1*, *SOD2*, and *WRN*, but none of the associations achieved conventional levels of statistical significance; the sample size was modest, and the genotyping platform did not cover the genome well by current standards. The advent of genome-wide association studies (GWAS) has successfully led to the discovery of novel genetic variants that have strong evidence for replication and that are outside of traditional candidate gene regions for several common diseases (23–29). The detection of novel genetic variants associated with longevity holds the promise to provide important insights to biologic pathways in the aging process and thus the potential to develop innovative strategies to promote a long and healthy life.

We conducted a meta-analysis of GWAS findings for longevity within an international consortium of four longitudinal community-based cohort studies that followed adults over many years. Longevity was defined as survival to age 90 years or older, and a comparison group was drawn from each cohort. Furthermore, we identified two independent cohorts of long-lived individuals, the Leiden Longevity Study cohort and the Danish 1905 cohort, to evaluate initial findings for the strongest allelic associations for longevity in a second discovery stage.

Methods

Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium

The Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium was convened to promote the discovery of new genomic loci involved in multiple complex traits in population-based follow-up studies using genome-wide association analysis (30). This meta-analysis used data from the CHARGE Consortium, which includes the Age, Gene/Environment Susceptibility-Reykjavik Study (AGES-Reykjavik) (31), the Cardiovascular Health Study (CHS) (32), FHS (33–36), and the Rotterdam Study (RS) (37).

The AGES-Reykjavik was funded by the National Institute on Aging (NIA) and was designed to examine genetic susceptibility and environmental interactions as risk factors for disease and disability in old age. Detailed phenotyping of the cardiovascular, neurocognitive, musculoskeletal, and body composition and metabolism was conducted in 5,764 men and women enrolled in 2002–2006 who were sampled from the 11,549 survivors of the AGES-Reykjavik of 30,000 men and women sampled from the 1907–1935 birth cohort (31). The CHS is a National Heart Lung and Blood Institute (NHLBI) contract-funded cohort study designed to evaluate risk factors for coronary heart disease (CHD) and stroke in older adults (32). Participants (n = 5,201) were recruited in 1989–1990, with an additional 687 minorities recruited in 1992–1993. The FHS is an NHLBI contract-funded cohort study initiated in 1948 to study determinants of CVD and other major illnesses. The original cohort comprised 5,209 men and women aged 28–62 years at enrollment who have undergone routine biennial examinations (33,34). In 1971, 5,124 offspring of the original cohort participants and offspring spouses, aged 5–70 years, were enrolled into the Framingham Offspring Study. Offspring participants have been examined approximately every 4–8 years (35,36). In the 1990s, DNA was obtained for genetic studies from surviving original cohort and offspring participants. The RS was planned and designed in the early 1990s as a longitudinal study investigating the incidence and progression of diseases in the elderly participants. From 1991 to 1995, all inhabitants of Ommoord, a district of Rotterdam in the Netherlands, who were aged 55 years or older, were invited to participate in this study (38). Of 10,275 eligible individuals, 7,983 agreed to participate (78%). The participants in the CHARGE studies are Caucasian by self-report. In each CHARGE study, population structure was assessed using principal components analysis, and outliers were removed. Any remaining within-study structure was adjusted for using appropriate methods (39). The details of each participating cohort study's genotyping platform, imputation algorithm, and quality control procedures used by each study are summarized in Supplementary Table 1. Each study was approved by the respective Institutional Review Board, and all participants provided consent.

Longevity and Comparison Group Definitions

In the present study, achievement of longevity was defined as reaching age 90 years or older, regardless of whether the participants were still living or had since died. Genotyped participants from these studies who died between the ages of 55 and 80 years were used as the comparison group. The comparison group was limited to deceased participants to ensure that no one in the comparison group could subsequently achieve longevity. The minimum age at death was set to match the minimum age at enrollment in the RS to promote age comparability of the comparison group across the four cohorts. The maximal age at death in the comparison group was set arbitrarily at age 80 years to include the majority of deaths, to maximize the overlap between birth cohorts, and to exclude those persons who survived far beyond average life expectancy for their respective birth cohort, that is persons who nearly reached longevity. Because of the timing of recruitment, DNA collection, and death, there was only partial overlap of the birth cohorts included in the comparison groups and the group of persons achieving longevity. Only Caucasian participants were included. Across the four studies, there were 1,836 persons who achieved longevity (144 from AGES-Reykjavik, 557 from CHS, 362 from FHS, and 773 from the RS), and the comparison group had 1,955 participants (122, 544, 355, and 934 participants from the AGES-Reykjavik, CHS, FHS, and RS, respectively). To facilitate comparison of results across studies, we imputed to 2.5 million SNPs using the HapMap Centre d'Etude du Polymorphisme Humain European Ancestry–genotyped samples as a reference. The effective sample size for all but one of the top SNPs was more than 80% of the full sample size of 3,791, indicating that the SNPs that were not directly genotyped were imputed well in most studies.

Second Discovery Stage Genotyping

Among the top 24 independent regions with the strongest associations for longevity in the four-study meta-analysis ($p < 10^{-4}$), we selected the 22 SNPs that had been tested in all four CHARGE cohorts in two additional Caucasian cohorts: the Leiden Longevity Study cohort and the Danish 1905 cohort. We excluded the two SNPs that could not be genotyped or imputed in all four CHARGE cohorts. Of the 22 SNPs selected for genotyping, 2 could not be genotyped and 4 did not pass quality control procedures; thus, 16 SNPs were analyzed in the second stage.

In the "Leiden Longevity Study" (7,40), a total of 950 long-lived proband siblings (mean age 94 years, range 89–104 years), 1,750 offspring (mean age 61 years, range 39–81 years), and 758 partners of offspring (mean age 60 years, range 36–79 years) were included. The additional genotyping of selected SNPs was undertaken in all 950 long-lived probands, and these were compared with the 744 partners of their offspring and an additional 680 blood bank donors

(60% men, mean age 31 years, range 18–40 years). All long-lived individuals and the comparison groups were from the Leiden area in the Netherlands and of European ancestry.

Participants in the "Danish 1905 Cohort Survey" are from the Danish 1905 birth cohort ascertained in 1998 when they were aged 92–93 years (41). Of the 3,600 participants alive from that cohort, 2,262 participants enrolled in the study. Participants underwent a home-based interview on health and lifestyle parameters, physical and cognitive tests, and collection of biologic material. The current genetic study comprises a total of 1,644 participants from this survey, mean age 93 years (range 92–93 years), 28% men. A comparison group included 2,007 Caucasians who were twins (one twin per pair) collected from all over Denmark, with a mean age of 57 years (range 46–68 years), 45% men.

Second Discovery Stage Genotyping Methods

Genotyping of the selected SNPs was performed using an iPLEX genotyping assay developed for use with the MassARRAY platform (Sequenom, Inc., San Diego, CA) (42). The iPLEX genotyping assay is based on mass spectrometry and enables genotyping of 25–36 custom SNPs on a sample in a single reaction. For the purposes of quality control, the system first automatically calls the genotypes and then generates cluster plots for all SNPs that are inspected individually by experienced technicians who check whether the plots show clear separation of the genotype clusters. There were two SNPs that did not pass quality control and two SNPs where no heterozygotes could not be detected; thus, lack of Hardy–Weinberg equilibrium was the quality control. Negative controls were included in the genotyping procedure (8 per 384-well plate), and importantly, 4% of samples were genotyped twice to confirm reproducibility (reproducibility was ≥99.7%).

Statistical Analysis

Using logistic regression, each imputed and observed HapMap SNP was tested for association with the longevity outcome using an additive genetic model adjusting for sex. The mean dosage of one of the alleles (a value between 0 and 2) was the predictor for imputed SNPs. The CHS additionally adjusted for field study site in the regression model, and the FHS used generalized estimating equations to account for familial correlations. We used the ratio of observed to expected variance in the imputed SNP genotype counts as a quality control metric for imputed SNPs (43). This ratio, multiplied by the sample size, is an estimate of the effective sample size. In the imputation software MaCH, this ratio is called r^2 as it is an estimate of the allelic correlation between the imputed genotypes and the true genotypes for the SNP. A total of 2,287,520 SNPs that had average minor allele frequency greater than 0.01 and were genotyped or imputed in all studies with variance ratio greater than 0.1 were meta-analyzed. 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 *df*. We calculated a meta-analysis odds ratio (OR) for each SNP using a fixed-effects model that combined logistic regression parameters and standard errors across the studies using inverse variance weights. The meta-analysis OR represents the increase in log-odds of surviving to age 90 years or older versus dying between ages 55 and 80 years for each additional copy of the minor allele of the SNP. SNP associations were considered to be significant on a genome-wide level at $p < 5 \times 10^{-8}$. The 16 SNPs in the second discovery phase effort were analyzed in the two study samples using an additive model. The results were added to the previous meta-analysis using a fixed-effects model as described earlier. Finally, using the top 24 results, we conducted a pathway analysis with the Database for Annotation, Visualization and Integrated Discovery (http://david.abcc.ncifcrf.gov/).

Results

Table 1 provides the characteristics of the persons achieving longevity and the comparison group in each of the four CHARGE discovery cohorts at the time of DNA collection. In line with the design of the study, persons achieving longevity were 10–20 years older than participants in the comparison group at baseline and were more likely to be women. Between 45% and 83% of those achieving longevity were still alive at the time that longevity status was ascertained. Among those who had died, the distributions of causes of death differed between those achieving longevity and the comparison group. Whereas 6%–12% of those achieving longevity died of cancer, more than 30% of the comparison group had death attributed to cancer. The prevalence of diabetes and a history of ever smoking were higher in the comparison group than in persons achieving longevity. The baseline prevalence of other cardiovascular risk factor levels showed substantial overlap between the two groups.

The genomic control inflation factor lambda (λ_{GC}) for each cohort was less than 1.05 (45). After meta-analysis, overall inflation of the meta-analysis *p* values was minor ($\lambda_{GC} = 1.034$; Figure 1). None of the SNP–longevity associations achieved the prespecified level of genome-wide significance of $p < 5 \times 10^{-8}$ (Figures 1 and 2). There were 273 SNP associations with meta-analysis $p < 10^{-4}$, and of these, 7 SNP associations had $p < 10^{-5}$ (Supplementary Table 2). Under the null hypothesis that there are no associations in the genome, we would expect $0.0001 \times \sim 2.3$ million = ~230 hits. Table 2 shows the top 24 independent SNPs associated with longevity along with the number of supporting SNPs (additional SNPs with linkage disequilibrium $r^2 > .80$ and $p < 10^{-4}$). Thus, for example, there were

Table 1. Characteristics of Longevity Cases and Comparison Group at DNA Collection

	CHS		Framingham Heart Study		Rotterdam Study		AGES-Reykjavik	
Characteristic, *M* (*SD*) or %	Survival Age to >90 y, *n* = 557	Comparison Group, *n* = 544	Survival Age to >90 y, *n* = 362	Comparison Group, *n* = 355	Survival Age to >90 y, *n* = 773	Comparison Group, *n* = 934	Survival Age to >90 y, *n* = 144	Comparison Group, *n* = 122
Age at DNA draw, y	79.6 (4.5)	69.5 (3.0)	87.3 (3.8)	66.5 (6.9)	83.7 (5.53)	66.5 (5.37)	88.0 (2.4)	73.8 (3.2)
Women, %	61	54	70	34	79	41	56	43
Alive, %	45	0	36	0	33	0	83	0
Cause of death*								
CVD, %	39	33	22	23	34	32	48	39
Cancer, %	10	40	9	45	6	39	12	38
Other, %	50	27	57	25	52	27	40	23
Unknown, %	0.3	0.2	12	6	7	2	0	0
Body mass index, kg/m^2	25.5 (3.9)	26.6 (5.2)	26.0 (4.1)	28.0 (5.5)	26.8 (3.81)	26.3 (3.75)	25.9 (4.0)	27.4 (4.7)
Ever smoker, %	40	70	54	81.0	29	43	49.3	80
Hypertension, %	57	53	68	75	40	40	83	80
Diabetes, %	8	20	8	22	6	8	8	11
Total cholesterol, mg/dL	210.5 (40.2)	212.2 (38.7)	198.8 (38.1)	(204.7 (47.1)	248 (49.4)	254 (46.8)	207.6 (44.3)	224.44 (42.7)

Notes: In the CHS, ever smoking was defined as having smoked more than 100 cigarettes or five packs during the participant's lifetime; hypertension was defined as a systolic blood pressure 140 mmHg or more or a diastolic blood pressure 90 mmHg or more or a history of hypertension and taking antihypertensive medication; diabetes was defined as fasting glucose more than 125 mg/dL or use of insulin or oral hypoglycemic medications. In Framingham Heart Study, ever smoking was defined as self-reported cigarette smoking of at least 1 cigarette/d for a year at any attended examination; total serum cholesterol was measured using an automated enzymatic procedure (44); hypertension was defined as blood pressure 140/90 mmHg or more or on antihypertensive medication; diabetes was defined as fasting blood glucose more than 125 mg/dL, a random blood glucose of more than 200 mg/dL, or use of insulin or oral hypoglycemic agents. In the Rotterdam Study, ever smoking was defined as self-reported ever smoking (cigarette, cigar, or pipe); hypertension was defined as systolic blood pressure 160 mmHg or more and/or diastolic blood pressure 100 mmHg or more and/or blood pressure–lowering medication with an indication for hypertension; total serum cholesterol was measured using an automated enzymatic procedure (40); diabetes was defined as self-reported diabetes at baseline. In the AGES-Reykjavik, ever smoking was defined as having smoked more than 100 cigarettes in one's lifetime; total serum cholesterol was measured using an automated enzymatic procedure (40); hypertension was defined as systolic blood pressure 140 mmHg or more, diastolic blood pressure 90 mmHg or more, use of antihypertensive medications, or self-report; diabetes was defined as fasting glucose more than 125 mg/dL, use of insulin or oral hypoglycemic medications, or self-report. AGES-Reykjavik = Age, Gene/Environment Susceptibility-Reykjavik Study; CHS = Cardiovascular Health Study; CVD = cardiovascular disease.

*As a proportion of all deaths for those in the survival to age 90 years or older group.

19 supporting SNPs on chromosome 15 in or near the longevity assurance homologue 3 (*LASS3*) gene, with the strongest association (OR = 0.79, $p = 1.2 \times 10^{-5}$) noted for rs8029244. The study-specific ORs for the 24 SNP associations shown in Table 2 were in the same direction and were of similar magnitude across the four cohorts (Figure 3; Supplementary Table 3).

Of the 24 strongest independent regions shown in Table 2, the 22 SNPs tested in all four CHARGE cohorts were selected for further evaluation, and 16 were successfully genotyped in the second stage cohorts. Only 1 of the 16 SNPs had a smaller *p* value after including the replication

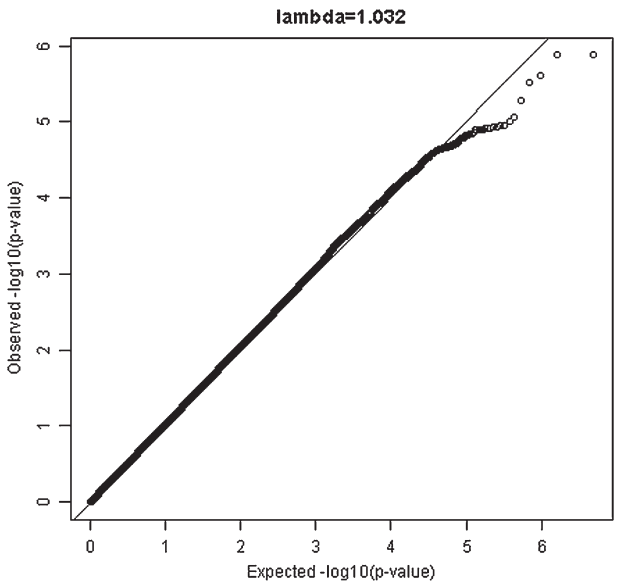


Figure 1. Quantile–quantile plot for the 2,287,520 single-nucleotide polymorphisms in the meta-analysis of survival to age 90 years or older.

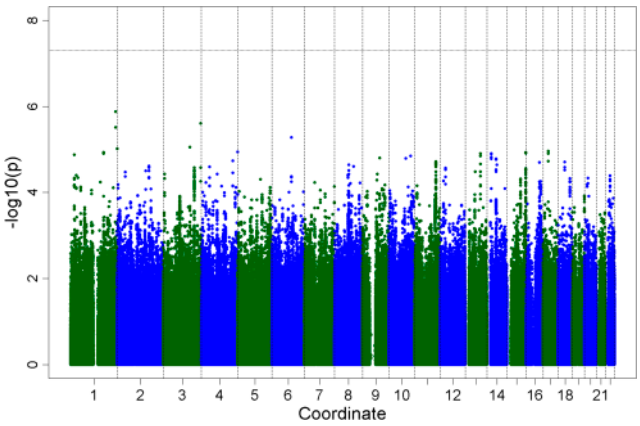


Figure 2. Plot of genome-wide association study for longevity meta-analysis (persons surviving to age ≥90 years, *n* = 1,836, and comparison group, *n* = 1,955) showing the $-\log_{10}$ (*p* values) based on the fixed-effects meta-analysis by chromosome. Line indicates threshold for genome-wide significance of 5×10^{-8}.

Table 2. Top 24 SNP Associations Ranked by *p* Value for Meta-analysis of Four Cohorts for Survival to Age 90 Years or Older (n = 1,836) Compared With Survival to Age 55–80 Years (n = 1,955)* and Second Discovery Stage Meta-analysis Results

Hit Number	Marker Name	Chromosome	Position	Closest Reference Gene	Distance (bp) From Closest Gene or Function	Alleles					Study Effect Direction	Number of Supporting SNPs	Effective Sample Size (proportion of total)	Second Discovery Stage Meta-analysis–CHARGE Plus Leiden and Denmark Cohorts		
						Minor	Major	MAF	OR	*p* Value				Study Effect Direction Leiden/Denmark	OR	*p* Value
1	rs4443878	1	238971041	*RGS7*	34,398	T	C	.04	0.41	1.3×10^{-6}	−−−−	10	2,435	++	0.83	.068
2	rs9825185	3	196415923	*C3orf21*	Intron	C	A	.13	1.44	2.5×10^{-6}	++++	24	0.64	−−	1.10	.045
3	rs954551	6	102886028	*GRIK2*	262,554	G	A	.25	0.77	5.3×10^{-6}	−−−−	55	0.88	†	—	—
4	rs7624691	3	138345769	*IL20RB*	133,159	C	T	.43	0.80	8.8×10^{-6}	−−−−	67	0.92	++	0.95	.092
5	rs10888267	1	246126046	*OR2W3*	Missense	C	T	.45	1.25	9.7×10^{-6}	++++	2	0.92	†	—	—
6	rs9972933	17	28589381	*ACCN1*	Intron	T	C	.23	0.77	1.1×10^{-5}	−−−−	30	0.91	+−	0.89	.003
7	rs2739532	4	190944424			C	G	.27	1.48	1.1×10^{-5}	?+?+	9	0.99	‡	—	—
8	rs8029244	15	98826098	*LASS3*	Intron	A	G	.49	0.79	1.2×10^{-5}	−−−−	108	0.37	+−	0.90	.002
9	rs16850255	1	175085164	*PAPPA2*	6,573	C	T	.21	0.75	1.2×10^{-5}	−−−−	81	0.86	++	0.92	.041
10	rs1543505	14	22432468	*REM2*	5,739	G	A	.28	1.27	1.3×10^{-5}	++++	23	0.83	−+	1.12	.001
11	rs7321904	13	80681190	*SPRY2*	868,103	T	C	.07	0.64	1.3×10^{-5}	−−−−	40	0.93	++	0.92	.179
12	rs17401847	1	20111053	*OTUD3*	971	G	A	.15	1.38	1.4×10^{-5}	++++	26	0.90	−+	1.12	.015
13	rs3124736	10	115487102	*CASP7*	6,450	A	G	.03	2.30	1.4×10^{-5}	+++?	0	0.89	‡	—	—
14	rs690232	9	92422258	*DIRAS2*	Intron	A	G	.30	1.27	1.6×10^{-5}	+++−	68	0.64	†	—	—
15	rs9664222	10	89328613	*MINPP1*	25,489	A	C	.21	0.77	1.6×10^{-5}	−−−−	55	0.93	−−	0.82	6.77×10^{-7}
16	rs11157721	14	49586464	*LOC196913*	33,655	T	C	.39	0.79	1.7×10^{-5}	−−−−	53	0.86	+−	0.90	.002
17	rs4690810	4	166500130	*SC4MOL*	16,456	C	T	.35	0.79	1.9×10^{-5}	−−−−	81	0.85	++	0.93	.044
18	rs11605096	11	113047320	*TMPRSS5*	16,162	A	C	.12	0.71	1.9×10^{-5}	−−−−	94	0.93	†	—	—
19	rs16972414	18	35709920	*PIK3C3*	2,079,276	G	A	.30	0.79	2.0×10^{-5}	−−−−	94	0.93	†	—	—
20	rs12935091	16	70082709	*ZNF19*	1,954	G	A	.07	0.62	2.0×10^{-5}	−−−−	46	0.94	−+	0.80	.002
21	rs210332	14	53262222	*BMP4*	223,982	C	T	.19	1.33	2.3×10^{-5}	++++	2	0.65	†	—	—
22	rs17369174	8	76128602	*CRISPLD1*	19,256	G	T	.10	0.69	2.3×10^{-5}	−−−−	83	0.79	+−	0.86	.014
23	rs6721003	2	166931758	*SCN7A*	38,026	A	G	.45	1.23	2.4×10^{-5}	++++	55	0.91	−+	1.09	.006
24	rs4734457	8	101573533	*ANKRD46*	28,644	A	C	.10	1.75	2.5×10^{-5}	++++	0	0.99	+−	1.10	.098

Notes: *p* Values are for the inverse variance–weighted meta-analysis. Distances to genes are given in base pairs. Position is for NCBI Build 36. ORs are for each additional minor allele. Number of supporting SNPs: The number of SNPs within 500 kb of the top SNP that are in LD with the top SNP in the HapMap CEU release 22 ($r^2 \geq .10$) and have association $p < .05$. For imputed and direction, study-specific information is presented in the order: Age, Gene/Environment Susceptibility-Reykjavik Study, Rotterdam Study, FHS, CHS, Leiden, and Denmark. Direction: + = minor allele increases odds of survival more than 90; − = minor allele decreases odds of survival; ? = not tested. CEU = Centre d'Etude du Polymorphisme Humain European Ancestry; CHARGE = Cohorts for Heart and Aging Research in Genomic Epidemiology; CHS = Cardiovascular Health Study; FHS = Framingham Heart Study; LD = linkage disequilibrium; MAF = minor allele frequency; OR = odds ratio; SNPs = single-nucleotide polymorphisms.

*For information on all SNP associations with $p < 10^{-4}$, see Supplementary Table 2.

†Genotyping requested, not completed.

‡Genotyping not requested.

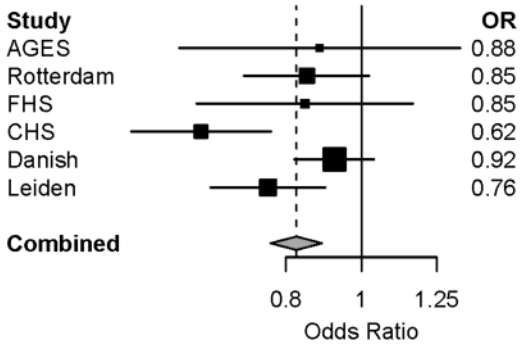


Figure 3. Study-specific odds ratios (ORs) and 95% confidence intervals for *MINPP1* (rs9664222) longevity association.

Note: AGES-Reykjavik = Age, Gene/Environment Susceptibility-Reykjavik Study; CHS = Cardiovascular Health Study; FHS = Framingham Heart Study.

studies in a joint meta-analysis, with the *p* value decreasing about 10-fold, from 1.61×10^{-5} to 6.77×10^{-7} and corresponding OR of 0.82. This SNP, rs9664222, is ~25 kb from the *MINPP1* gene (Figure 4). In the CHARGE analysis, the minor allele was associated with a lower odds of survival past age 90 (OR = 0.77). The Leiden study yielded a similar effect estimate (OR = 0.76, $p = .0014$), whereas the Danish study showed a nonsignificant trend in the same direction (OR = 0.92, $p = .19$). Findings for the other SNPs were inconsistent in direction of association such that the meta-analysis *p* values increased with inclusion of the second stage cohorts (Supplementary Table 3). Pathway analysis did not reveal significant findings in the top associations, though some groupings were biologically plausible.

Discussion

The CHARGE Consortium collaboration allowed us to conduct a meta-analysis of GWAS for longevity in a sample of long-lived individuals and a corresponding comparison group derived from the same longitudinal community-based cohort studies. Although none of the SNP associations for longevity in the first discovery phase achieved prespecified level of genome-wide significance, a polymorphism associated with the *MINPP1* genes was among the strongest associations observed in our sample, with effect sizes that were similar within the four cohorts. The finding related to the *MINPP1* gene was strengthened after including two additional cohorts in a second discovery phase but did not reach genome-wide significance. Among the top 10 associations in the initial meta-analysis, additional SNP associations of potential interest in longevity include SNPs in or near *LASS3*, *ACCN1*, *IL20RB*, and *PAPPA2*. These SNPs

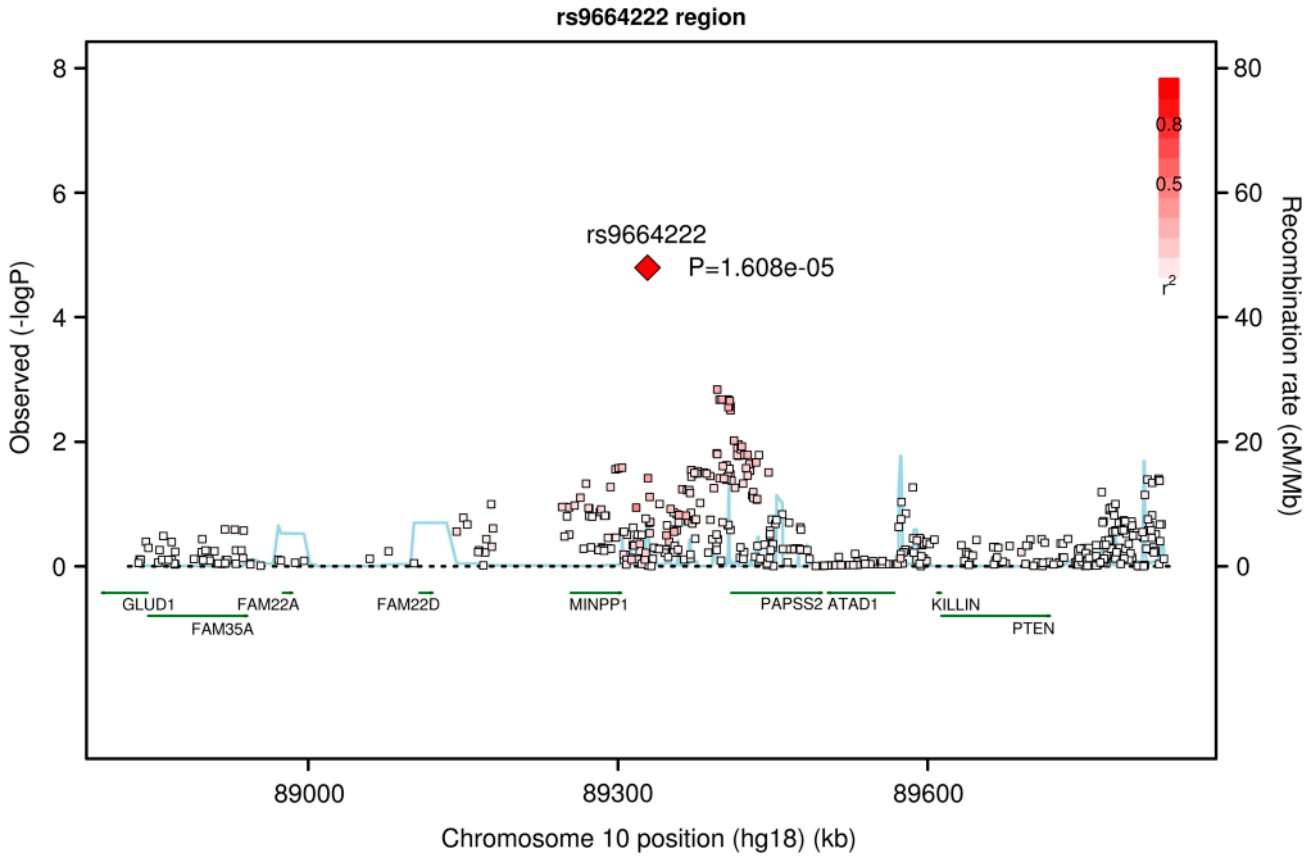


Figure 4. Regional plot for rs9664222 near *MINPP1*.

are near genes that have not previously been reported to be associated with longevity in human populations but are interesting because these genes are conserved in basic biologic pathways.

The *MINPP1* gene codes multiple inositol polyphosphate phosphatases, which are compartmentalized to the endoplasmic reticulum lumen. *MINPP1*-deficient mice have no obvious defects, though targeted deletion in vitro is associated with slowed cellular proliferation (46). There is no evidence that this SNP is functional; furthermore, its distance from the gene shows that it is not in strong linkage disequilibrium with SNPs in *MINPP1* (47). However, it is well known that important regulatory elements are found outside of genes. This SNP is within 50 kb of two copy number variants. The finding of an SNP near a gene regulating proliferation is intriguing because of the higher rate of cancer death in the comparison group.

The initial finding in the *LASS3* gene region was of interest because of the historical association of its homologue with longevity in yeast (46). The *LASS* gene family contains a group of highly conserved genes that are found in all eukaryotic species. *LASS* isoforms are mammalian homologues of the yeast longevity assurance gene 1, which encodes a protein that regulates life span (48). The strongest association was noted for rs8029244; this SNP is in the intronic enhancer region of the *LASS3* gene. *LASS3* is a member of the ceramide synthase family, which is important in sphingolipid metabolism, cell differentiation, cell cycling, and apoptosis (46). *LASS3* may be involved in sphingolipid synthesis or its regulation (49).

IL20RB, interleukin 20 receptor beta IL-20, plays a role in skin inflammation and the development of hematopoietic cells (50) and is of interest because of the strong associations of inflammation with the aging process (51). IL-20 is a pleiotropic cytokine with potent inflammatory, angiogenic, and chemoattractive characteristics and is involved in inflammatory diseases, such as psoriasis, atherosclerosis, and rheumatoid arthritis (50). The *ACCN1* gene encodes amiloride-sensitive sodium channels with two hydrophobic transmembrane regions and a large extracellular loop, which has many cysteine residues with conserved spacing (52,53). The member encoded by this gene may play a role in neurotransmission. *ACCN1* was found to be associated with multiple sclerosis (54). Pregnancy-associated plasma protein A2 (*PAPPA2*) is a metalloproteinase regulating local insulin-like growth factor pathway action (55). Genetic deletion extends life span in the mouse by 30%–40% (56) and is characterized by delay in thymic involution (57) and low rates of tumor incidence (56). Although the associations reported here did not reach the a priori specified level of significance, the findings are important to report so that they can be replicated in studies without whole-genome genotyping and compared with future studies, such as in centenarian studies and family studies of longevity. Effect size estimates noted here support the likelihood that longevity is a complex process, in that there were no variants with large effects, supporting the hypothesis that there may be many genes with small effects that contribute to longevity.

The strengths of this study include the community-based prospective design and the long-term follow-up of these cohorts. In all cases, vital status was confirmed using death certificates and hospital records. Another strength was our ability to use controls that were equally well characterized and were drawn from within the same cohorts. The number of long-lived individuals reported here is very large relative to other studies in the literature, allowing greater ability to identify SNPs with small effects. The cohorts were relatively homogeneous with respect to ancestry, limited to Caucasians of European decent. Our top associations were homogeneous across cohorts. Screening for latent population substructure also supported ethnic homogeneity. Thus, the findings reported are less likely to be due to population stratification.

There are important aspects of the study that need to be kept in mind when interpreting the results. The differences in causes of death in the longevous individuals versus the comparison groups are expected as death from cancer tends to occur earlier in life than death from heart disease or dementia. Many of the long-lived people are still alive and we do not yet know what their ultimate cause of death will be, but it is likely that cancer will be underrepresented among persons achieving longevity. Power remains a limitation. Thus, future GWAS aiming to identify variants for this phenotype will have to consider small effect sizes and target a sample size larger than our nearly 2,000 long-lived persons. DNA collection in cohort studies is a recent enough phenomenon that relatively few cohort members who had DNA collected have had the opportunity to survive to age 90 years. Continuous study of these and other similarly designed cohorts will allow us to extend this study to larger numbers and to older ages.

In our case comparison analysis, we attempted to account for birth cohort, but the overlap between birth year of the comparison group and of the long-lived participants was limited. Further follow-up of these cohorts is needed to increase our ability to examine potential birth cohort effects. The study design of the cohorts examined in the second stage was different from the initial four-study CHARGE meta-analysis in that the comparison groups were derived from younger participants, living and deceased, who were not from the same cohort as the individuals achieving longevity. Certainly, there are important environmental factors that would be necessary for the fulfillment of the genetic potential for longevity. Heterogeneity in environmental exposures and gene–environment interactions require further study. Finally, these results cannot be extended to populations of other ancestry.

In conclusion, this meta-analysis of GWAS data for longevity from four large cohorts and two additional

cohorts has implicated several genes involved in conserved basic mechanisms of cellular function. Analysis of more extreme survival phenotypes such as centenarians, additional follow-up to increase sample size in these cohorts for this phenotype, or evaluation of more specific phenotypes such as disease-free survival may support and refine these initial findings.

Funding

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, and N01-HC-45133 and grant numbers U01 HL080295 and R01 HL087652 from the NHLBI and R01 AG023629 from the NIA, 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 http://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. The FHS phenotype–genotype analyses were supported by the NIA grant numbers R01AG029451, R01AG028321, R01AG033193, and R01 AG031287. The FHS of the NHLBI of the National Institutes of Health (NIH) and Boston University School of Medicine were supported by the NHLBI's FHS contract number N01-HC-25195 and its contract with Affymetrix, Inc. for genotyping services (contract number N02-HL-6-4278). Analyses reflect intellectual input and resource development from the FHS investigators participating in the SNP Health Association Resource 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. D.P.K.'s effort was supported by a grant from the National Institute of Arthritis, Musculoskeletal and Skin Diseases and the NIA (R01 AR/AG 41398). The RS is supported by the Erasmus Medical Center and Erasmus University Rotterdam; the Netherlands Organisation for Scientific Research (NWO); the Netherlands Organisation for Health Research and Development (ZonMw); the Research Institute for Diseases in the Elderly; the Ministry of Education, Culture and Science; the Ministry of Health, Welfare and Sports; and the European Commission (DG XII). The genetic analyses were supported by the Netherlands Organization for Scientific Research grant number 175.01.2005.011. This study was supported by the Netherlands Genomics Initiative-Sponsored Netherlands Consortium for Healthy Aging (nr 050-060-810). The work of S.W. was supported by Netspar. The AGES-Reykjavik is funded by NIH contract number 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 Intramural Research Program Laboratory of Neurogenetics. The Leiden Longevity Study was supported by the Innovation Oriented research Program on Genomics, the Centre for Medical Systems Biology, and the Netherlands Consortium for Healthy Aging in the framework of the Netherlands Genomics Initiative, the Netherlands Organization for Scientific Research, the Netherlands Organization for Health Research and Development (ZonMw), the Danish National Programme for Research Infrastructure 2007 from the Danish Agency for Science Technology and Innovation, and Unilever PLC, UK. The Longevity Consortium, funded by the NIA, grant number U19 AG023122, provided administrative resources to CHARGE investigators for this phenotype as well as scientific opportunity funds to conduct follow-up genotyping. The Danish 1905 Cohort Survey was supported by grants from the Danish National Research Foundation and NIH/NIA grant P01 AG08761. The Danish Aging Research Center is supported by a grant from the Velux Foundation. The Atherosclerosis Risk in Communities Study is carried out as a collaborative study supported by NHLBI contract numbers N01-HC-55015, N01-HC-55016, N01-HC-55018, N01-HC-55019, N01-HC-55020, N01-HC-55021, N01-HC-55022, R01HL087641, R01HL59367, and R01HL086694; National Human Genome Research Institute contract number U01HG004402; and NIH contract number HHSN268200625226C. Infrastructure was partly supported by grant number UL1RR025005, a component of the NIH and NIH Roadmap for Medical Research.

Conflict of Interest

The content is solely the responsibility of the authors and does not necessarily represent the views of the NIA, NHLBI, the National Institute of Neurological Disorders and Stroke, or NIH.

Supplementary Material

Supplementary material can be found at: http://biomed.gerontologyjournals.org/

Acknowledgments

We are indebted to the participants and staff in the AGES-Reykjavik, the CHS, the FHS, and the RS for their important contributions. We acknowledge the NHLBI, which has made the Framingham SNP Health Association Resource project possible.

References

1. Perls T, Shea-Drinkwater M, Bowen-Flynn J, et al. Exceptional familial clustering for extreme longevity in humans. *J Am Geriatr Soc*. 2000;48(11):1483–1485.
2. Perls T, Kohler IV, Andersen S, et al. Survival of parents and siblings of supercentenarians. *J Gerontol A Biol Sci Med Sci*. 2007;62(9):1028–1034.
3. Perls T, Terry D. Genetics of exceptional longevity. *Exp Gerontol*. 2003;38(7):725–730.
4. Hjelmborg JV, Iachine I, Skytthe A, et al. Genetic influence on human lifespan and longevity. *Hum Genet*. 2006;119(3):312–321.
5. Herskind AM, McGue M, Holm NV, Sorensen TI, Harvald B, Vaupel JW. The heritability of human longevity: a population-based study of 2872 Danish twin pairs born 1870-1900. *Hum Genet*. 1996;97(3):319–323.
6. Iachine IA, Holm NV, Harris JR, et al. How heritable is individual susceptibility to death? The results of an analysis of survival data on Danish, Swedish and Finnish twins. *Twin Res*. 1998;1(4):196–205.
7. Schoenmaker M, de Craen AJM, de Meijer PHEM, et al. Evidence of genetic enrichment for exceptional survival using a family approach: the Leiden Longevity Study. *Eur J Hum Genet*. 2005;14(1):79–84.
8. Butler RN, Austad SN, Barzilai N, et al. Longevity genes: from primitive organisms to humans. *J Gerontol A Biol Sci Med Sci*. 2003;58(7):581–584.
9. Browner WS, Kahn AJ, Ziv E, et al. The genetics of human longevity. *Am J Med*. 2004;117(11):851–860.
10. Franceschi C, Olivieri F, Marchegiani F, et al. Genes involved in immune response/inflammation, IGF1/insulin pathway and response to oxidative stress play a major role in the genetics of human longevity: the lesson of centenarians. *Mech Ageing Dev*. 2005;126(2):351–361.
11. van Heemst D, Beekman M, Mooijaart SP, et al. Reduced insulin/IGF-1 signalling and human longevity. *Aging Cell*. 2005;4(2):79–85.
12. Holzenberger M, Dupont J, Ducos B, et al. IGF-1 receptor regulates lifespan and resistance to oxidative stress in mice. *Nature*. 2003;421(6919):182–187.
13. Matsuoka S, Ballif BA, Smogorzewska A, et al. ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. *Science*. 2007;316(5828):1160–1166.
14. Reiner AP, Diehr P, Browner WS, et al. Common promoter polymorphisms of inflammation and thrombosis genes and longevity in older adults: the Cardiovascular Health Study. *Atherosclerosis*. 2005;181(1):175–183.
15. Reiner AP, Carlson CS, Jenny NS, et al. USF1 gene variants, cardiovascular risk, and mortality in European-Americans. Analysis of two U.S. cohort studies. *Arterioscler Thromb Vasc Biol*. 2007;27:2736–2742.
16. Walston JD, Fallin MD, Cushman M, et al. IL-6 gene variation is associated with IL-6 and C-reactive protein levels but not cardiovascular outcomes in the Cardiovascular Health Study. *Hum Genet*. 2007;122:485–494.

17. Arking DE, Atzmon G, Arking A, Barzilai N, Dietz HC. Association between a functional variant of the KLOTHO gene and high-density lipoprotein cholesterol, blood pressure, stroke, and longevity. *Circ Res*. 2005;96(4):412–418.
18. Atzmon G, Rincon M, Schechter CB, et al. Lipoprotein genotype and conserved pathway for exceptional longevity in humans. *PLoS Biol*. 2006;4(4):e113.
19. Barzilai N, Atzmon G, Schechter C, et al. Unique lipoprotein phenotype and genotype associated with exceptional longevity. *JAMA*. 2003;290(15):2030–2040.
20. Puca AA, Daly MJ, Brewster SJ, et al. A genome-wide scan for linkage to human exceptional longevity identifies a locus on chromosome 4. *Proc Natl Acad Sci U S A*. 2001;98(18):10505–10508.
21. Reed T, Dick DM, Uniacke SK, Foroud T, Nichols WC. Genome-wide scan for a healthy aging phenotype provides support for a locus near D4S1564 promoting healthy aging. *J Gerontol A Biol Sci Med Sci*. 2004;59(3):227–232.
22. Lunetta KL, D'Agostino RB, Sr, Karasik D, et al. Genetic correlates of longevity and selected age-related phenotypes: a genome-wide association study in the Framingham Study. *BMC Med Genet*. 2007;8(suppl 1):S13.
23. Easton DF, Pooley KA, Dunning AM, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature*. 2007;447: 1087–1093.
24. Hunter DJ, Kraft P, Jacobs KB, et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat Genet*. 2007;39: 870–874.
25. Gudmundsson J, Sulem P, Manolescu A, et al. Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nat Genet*. 2007;39(5):631–637.
26. Yeager M, Orr N, Hayes RB, et al. Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat Genet*. 2007;39(5):645–649.
27. Thomas G, Jacobs KB, Yeager M, et al. Multiple loci identified in a genome-wide association study of prostate cancer. *Nat Genet*. 2008;40(3):310–315.
28. Saxena R, Voight BF, Lyssenko V, et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science*. 2007;316: 1331–1336.
29. Helgadottir A, Thorleifsson G, Manolescu A, et al. A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science*. 2007;316: 1491–1493.
30. Psaty BM, O'Donnell CJ, Gudnason V, et al. Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium: design of prospective meta-analyses of genome-wide association studies from five cohorts. *Circ: Cardiovasc Genet*. 2009;2: 73–80.
31. Harris TB, Launer LJ, Eiriksdottir G, et al. Age, Gene/Environment Susceptibility-Reykjavik Study: multidisciplinary applied phenomics. *Am J Epidemiol*. 2007;165(9):1076–1087.
32. Fried LP, Borhani NO, Enright P, et al. The Cardiovascular Health Study: design and rationale. *Ann Epidemiol*. 1991;1(3):263–276.
33. Dawber TR, Meadors GF, Moore FE Jr. Epidemiological approaches to heart disease: the Framingham Study. *Am J Public Health*. 1951; 41(3):279–281.
34. Dawber TR, Kannel W, Lyell L. An approach to longitudinal studies in a community: the Framingham Study. *Ann N Y Acad Sci*. 1963;107: 539–556.
35. Feinleib M, Kannel WB, Garrison RJ, McNamara PM, Castelli WP. The Framingham Offspring Study. Design and preliminary data. *Prev Med*. 1975;4(4):518–525.
36. Kannel W, 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(3):281–290.
37. Hofman A, Breteler MM, van Duijn CM, et al. The Rotterdam Study: objectives and design update. *Eur J Epidemiol*. 2007;22(11):819–829.
38. Hofman A, Grobbee DE, de Jong PT, van den Ouweland FA. Determinants of disease and disability in the elderly: the Rotterdam Elderly Study. *Eur J Epidemiol*. 1991;7(4):403–422.
39. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006;38(8):904–909.
40. van Gent CM, van der Voort HA, de Bruyn AM, Klein F. Cholesterol determinations. A comparative study of methods with special reference to enzymatic procedures. *Clin Chim Acta*. 1977;75(2):243–251.
41. Westendorp RGJ, van Heemst D, Rozing M, et al. Nonagenarian siblings and their offspring display lower risk of mortality and morbidity than sporadic Nonagenarians: the Leiden Longevity Study. *J Am Geriatr Soc*. 2009;57: 1634–1637.
42. Nybo H, Petersen HC, Gaist D, et al. Predictors of mortality in 2,249 nonagenarians—the Danish 1905-Cohort Survey. *J Am Geriatr Soc*. 2003;51(10):1365–1373.
43. Oeth P, Beaulieu M, Park C, et al. iPLEX assay: increased plexing efficiency and flexibility for Mass ARRAY system through single base primer extension with mass-modified terminators. www.sequenom.com. http://jmgroup.pl/kawaska/download/iPLEX%20Application%20note.pdf. Accessed November 10, 2006. Application note: 1–12.
44. de Bakker PI, Ferreira MA, Jia X, Neale BM, Raychaudhuri S, Voight BF. *Practical aspects of imputation-driven meta-analysis of genome-wide association studies*. 2008;17: R122–R128.
45. Devlin B, Roeder K. *Genomic control for association studies*. 1999;55: 997–1004.
46. Chi H, Yang X, Kingsley PD, et al. Targeted deletion of Minpp1 provides new insight into the activity of multiple inositol polyphosphate phosphatase in vivo. *Mol Cell Biol*. 2000;20(17):6496–6507.
47. Ioannidis JPA, Thomas G, Daly MJ. Validating, augmenting and refining genome-wide association signals. *Nat Rev Genet*. 2009;10(5):318–329.
48. Teufel A, Maass T, Galle PR, Malik N. The longevity assurance homologue of yeast lag1 (Lass) gene family (review). *Int J Mol Med*. 2009;23(2):135–140.
49. Mizutani Y, Kihara A, Igarashi Y. Mammalian Lass6 and its related family members regulate synthesis of specific ceramides. *Biochem J*. 2005;390(Pt 1):263–271.
50. Wei CC, Hsu YH, Li HH, et al. IL-20: biological functions and clinical implications. *J Biomed Sci*. 2006;13(5):601–612.
51. van den Biggelaar AHJ, Huizinga TWJ, de Craen AJM, et al. Impaired innate immunity predicts frailty in old age. The Leiden 85-plus study. *Exper Gerontol*. 2004;39(9):1407–1414.
52. Saugstad JA, Roberts JA, Dong J, Zeitouni S, Evans RJ. Analysis of the membrane topology of the acid-sensing ion channel 2a. *J Biol Chem*. 2004;279(53):55514–55519.
53. Waldmann R, Champigny G, Voilley N, Lauritzen I, Lazdunski M. The mammalian degenerin MDEG, an amiloride-sensitive cation channel activated by mutations causing neurodegeneration in Caenorhabditis elegans. *J Biol Chem*. 1996;271(18):10433–10436.
54. Bernardinelli L, Murgia SB, Bitti PP, et al. Association between the ACCN1 gene and multiple sclerosis in Central East Sardinia. *PLoS One*. 2007;2(5):e480.
55. Page NM, Butlin DJ, Lomthaisong K, Lowry PJ. The characterization of pregnancy associated plasma protein-E and the identification of an alternative splice variant 3. *Placenta*. 2001;22(8–9):681–687.
56. Conover CA, Bale LK. Loss of pregnancy-associated plasma protein A extends lifespan in mice 21. *Aging Cell*. 2007;6(5):727–729.
57. Vallejo AN, Michel JJ, Bale LK, Lemster BH, Borghesi L, Conover CA. Resistance to age-dependent thymic atrophy in long-lived mice that are deficient in pregnancy-associated plasma protein A. *Proc Natl Acad Sci*. 2009;106(27):11252–11257.

Journal of Health Communication, 15:224–225, 2010

ISSN: 1081-0730 print/1087-0415 online
DOI: 10.1080/10810730.2010.499993

Routledge
Taylor & Francis Group

Crossword Puzzle: A Decoding Challenge

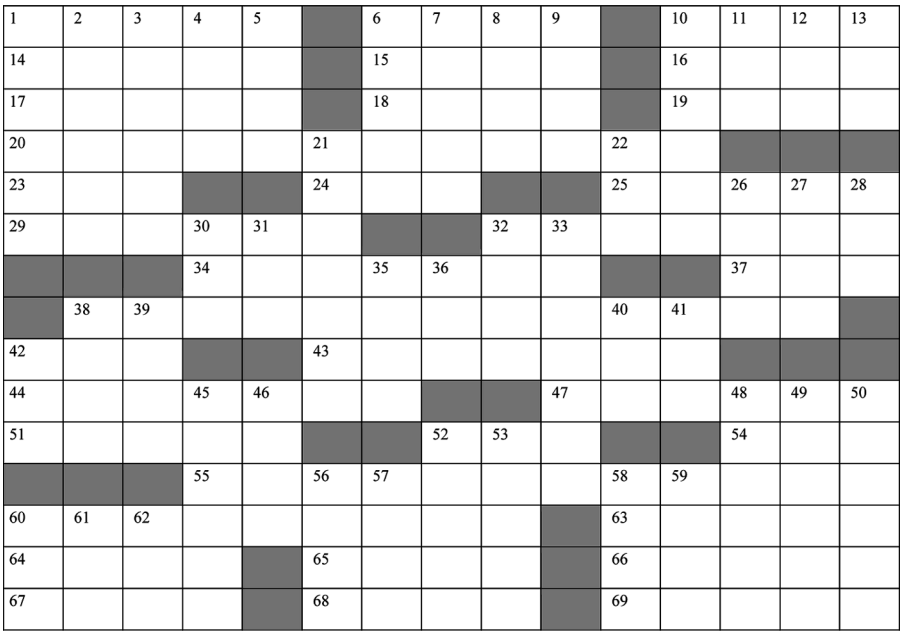
MICHAEL K. PAASCHE-ORLOW AND WENDY GRAY

Department of Medicine, Boston University School of Medicine, Boston, Massachusetts, USA

MARYA J. COHEN

Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA

The first person to submit a complete correct puzzle will be honored and receive a prize at the Health Literacy Annual Research Conference, October 18–19, 2010 (www.bumc.bu.edu/healthliteracyconference/). Submit to mpo@bu.edu or by fax to (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. The vowels
6. Start of Frederick Douglass quote, "____free."
10. Nimble
14. Rioja and tempranillo in Catalan
15. Heart and _____
16. Popular choice in the cookie aisle
17. Stared at
18. Of all sorts
19. Hosp. unit for children
20. *Letters To Those* _____, Paulo Friere
23. iPhone™ network
24. Before
25. Diadem
29. Destroying a cell membrane
32. Braced
34. What the card player did
37. Red Book author
38. "Outside of a dog a book is man's best friend.
    Inside of a dog it's ____."
    Groucho Marx
42. Slip up
43. "_____ free." Frederick Douglass quote, Part 4
44. "_____ free." Frederick Douglass quote, Part 3
47. Madden
51. Stinging ant
52. Maj. intest. blood supplier
54. Quagmire
55. "____, tomorrow a leader."
    Margaret Fuller
60. "_____ free." Frederick Douglass quote, Part 6
63. Finely-tuned or sharp
64. Iniquity
65. Spring in Jerusalem
66. Heats
67. "_____ free." Frederick Douglass quote, Part 5
68. Rasp
69. Vacation destinations

Down

1. Confirmation
2. Days in ark + days on mountain
3. Where diamonds are created
4. Expressed admiration
5. Meat recall grp.
6. Willow used in basketry
7. 48 HRS lead cop
8. Fruit nickname
9. Lioness that was born free
10. She had a choice
11. Distributor for "This American Life"
12. Horror film that has a sequel
13. "_____ free." Frederick Douglass quote, Part 2
21. Renaissance composer
22. Dementia Pugalistica alt.
26. ____ *mater*
27. "Reach Out and ____"
28. Difficulty
30. ____ Jima
31. Flanders or Mandingo
32. Daily musical select.
33. Type of molars
35. Palm phone
36. RTS the ___ is inaccurate according to USPS
38. Type of hugger
39. Type of presentation
40. Paul or Artest
41. Continent abbrev.
42. Moose
45. Restate
46. Wikipedia auth. rule
48. Move arm away from body
49. Chin beard
50. Exit
52. Field's multiple personality role
53. Intoxicating queen of Irish mythology
56. Stone or tone
57. Taro root dish
58. Measures of corn
59. Type of berry
60. Not many
61. Duct, prefix
62. What Meril and Avril have in common

Journal of Health Communication, 15:5–8, 2010

ISSN: 1081-0730 print/1087-0415 online
DOI: 10.1080/10810730.2010.499995

Routledge
Taylor & Francis Group

Introduction

The Evolving Field of Health Literacy Research

MICHAEL K. PAASCHE-ORLOW

Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, Boston, Massachusetts, USA

ELIZABETH A. H. WILSON

Division of General Internal Medicine, Feinberg School of Medicine, Northwestern University, Chicago, Illinois, USA

LAUREN McCORMACK

Health Communication Program, RTI International, Research Triangle Park, North Carolina, USA

We are quite pleased to present the current special issue on Health Literacy of the *Journal of Health Communication*. We hope readers will see this issue as a marker of the Journal's long-term interest and commitment to health literacy research and as an exhibit of the rapidly evolving field of health literacy. This issue presents findings from the first Health Literacy Annual Research Conference (HARC) which took place in October 2009 at the National Academy of Sciences Building, Washington, DC. HARC is an interdisciplinary meeting for investigators dedicated to health literacy research. Our aim is to attract a full range of investigators engaged in health literacy research including those involved in a broad array of public health, health services, epidemiology, translational, and interventional research activities.

The Health Literacy Annual Research Conference (HARC) meeting was coordinated with the assistance of Rose Martinez on behalf of the Institute of Medicine Roundtable on Health Literacy and sponsored by a grant from the National Center on Minority Health and Health Disparities (NCMHD), Project Officer—Robert Netty, with additional significant support from the Agency for Healthcare Research and Quality (AHRQ), Project Officer—Cindy Brach. Additional coordination and support was provided by Helen Meissner, of the Office of Behavioral and Social Sciences Research (OBSSR) as well as the National Institute of Biomedical Imaging & Bioengineering (NIBIB), Project Officer—John Haller, and the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), Project Officer—Lynne Haverkos. Principal Investigator Michael Paasche-Orlow, 1 R13 MD003392. This Special Issue was supported by generous funding from RTI International, the Health Literacy and Learning Program, Division of General Internal Medicine, Feinberg School of Medicine at Northwestern University, the George Washington University Center for Global Health, and the Section of General Internal Medicine, Boston University School of Medicine.

Address correspondence to Michael K. Paasche-Orlow, 801 Massachusetts Avenue, 2nd Floor, Boston, MA 02118, USA. E-mail: mpo@bu.edu

Establishing an interdisciplinary research home for health literacy investigators can accelerate: (1) professional development, (2) advancement of the science of health literacy research, and (3) promotion of interdisciplinary research.

The HARC conference took place over 2 full days in October 2009 and highlighted areas of important research advancement as well as key lacunae in the field. A keynote address by David Baker, MD, MPH examined the role health literacy in patient education, and a keynote address by Anne Beal, MD, MPH, focused on the role of health literacy in health disparities. The conference also included four panels of invited speakers dealing with: (1) measurement; (2) health literacy and verbal interactions; (3) health information technology interventions; and (4) organizational assessment and change. Current gaps in the research were examined by invited speakers and in breakout sessions relating to public health approaches to health literacy, health disparities and health literacy, and health information technology. An additional 62 posters and 12 oral abstracts were presented, making this clearly the largest set of Health Literacy research presentations in any single meeting to date. The meeting was attended by an interdisciplinary array of investigators representing the strong majority of health literacy investigators in the United States.

HARC II will take place October 18–19, 2010, in Bethesda, Maryland (http://www.bumc.bu.edu/healthliteracyconference/).

While many important unanswered questions loom, the articles in this issue highlight the blossoming nature of health literacy research. In her commentary, Health Literacy: The Second Decade for Distinction, Parker chronicles the major landmarks of the last decade. A Pubmed exercise tells a similar story. Using the Health Literacy topic specific query under the Pubmed tools tab (see: http://www.nlm.nih.gov/services/health_literacy.html) reveals the tremendous growth of research in this field: in the 5 years between 1986 and 1990 there are 129 references in Pubmed; in the years 1991 to 1995 this search yields 306 references; between 1996 and 2000 there are 307 references; from 2001 to 2005 there are 602 references from this search; and in the current interval, between 2006 and 6/6/2010, there are already 1576 references returned by this search (see Figure 1). As striking as this may seem, it

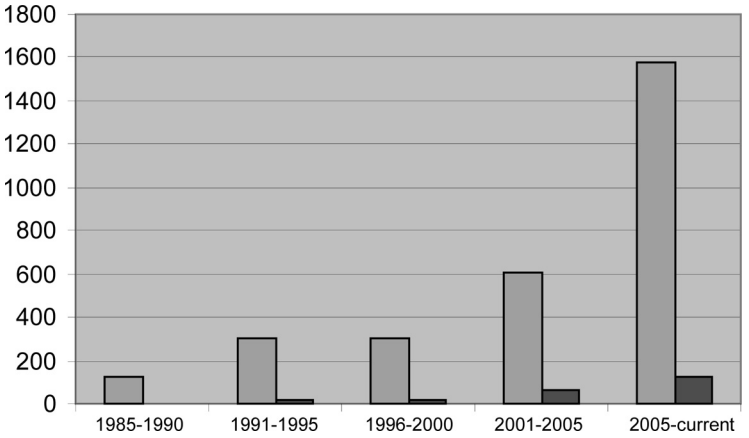


Figure 1. Health literacy citations in Pubmed. Lighter bars represent all citations identified by Health Literacy topic specific query. Darker bars represent the subset of these citations identified as trials, clinical trials, or randomized controlled trials.

is also important to note that this collection of references is dominated by observational research, indeed, fewer than 8% of these citations are tagged as trials.

An ongoing research area in the field relates to health literacy measurement. McCormack et al. presents pilot data from a new skills-based instrument, Yost and colleagues present data on the acceptability of a talking touchscreen assessment tool, and Gazmararian et al. describes a tool developed to assess the health literacy environment of health plans (McCormack et al.; Yost et al.; Gazmararian et al.). All such efforts inherently reflect underlying definitional issues, which are also the focus of a Commentary by Berkman et al.

This special issue also provides evidence about health literacy in an array of contexts including advanced care planning (Sudore et al.), informed consent for research among Spanish speakers (Cortés et al.), two articles about cancer screening (Wilson et al. and Mazor et al.), and the relationship between health literacy and the quality of communication in health care organizations (Wynia et al.). Self-efficacy is shown to link literacy and numeracy in glycemic control (Osborn et al.) and a new instrument of patient self-efficacy regarding communication with clinical encounters (Clayman et al.). Despite evidence of a large 'digital divide' by level of health literacy in accessing and using a patient-portal (Sarkar et al.), it appears that such barriers are surmountable, as Bickmore et al. present evidence from two trials of an 'embodied conversational agent' technology interface that is usable by people with limited literacy (Bickmore et al.).

We have also included a Commentary on the role of health literacy research for the elimination of health disparities (Paasche-Orlow & Wolf) and a Commentary that situates the field of health literacy within a public health paradigm (Baur). In addition, we have included an international policy paper that summarizes the health literacy objectives of the United Nations Millennium Development Goals (UN Economic and Social Council [ECOSOC]).

Taken together, the special issue reflects the important role that health literacy plays in many aspects of health communication. The Department of Health and Human Services recently outlined a National Action Plan to Improve Health Literacy (http://www.health.gov/communication/HLActionPlan/). Clearly, improving health literacy is a widespread goal. However, it is a means to end. Ultimately, improved health literacy can lead to clearer communication, better informed decisions, and the delivery of quality health care services. To achieve these objectives, we need to continue to study how to measure health literacy, which interventions can improve health literacy levels, and the relationships between health literacy and health outcomes.

References

Baur, C. (2010). New directions in research on public health and health literacy. *Journal of Health Communication*, *15*(S2), 42–50.

Benjamin, R. (2010). Health literacy improvement as a national priority. *Journal of Health Communication*, *15*(S2), 1–3.

Berkman, N. D., Davis, T. C., & McCormack, L. A. (2010). Health literacy: What is it? *Journal of Health Communication*, *15*(S2), 9–19.

Bickmore, T. W., Pfeifer, L. M., Byron, D., Forsythe, S., Henault, L. E., Jack, B. W., et al. (2010). Usability of conversational agents by patients with inadequate health literacy: Evidence from two clinical trials. *Journal of Health Communication*, *15*(S2), 197–210.

Clayman, M. L., Pandit, A. U., Bergeron, A. R., Cameron, K. A., Ross, E., & Wolf, M. S. (2010). Ask, understand, remember: A brief measure of patient communication self-efficacy within clinical encounters. *Journal of Health Communication*, *15*(S2), 72–79.

Cortés, D., Drainoni, M.-L., Henault, L. E., & Paasche-Orlow, M. (2010). How to achieve informed consent for research from Spanish-speaking individuals with low literacy: A qualitative report. *Journal of Health Communication*, *15*(S2), 172–182.

Gazmararian, J. A., Beditz, K., Pisano, S., & Carreón, R. (2010). The development of a health literacy assessment tool for health plans. *Journal of Health Communication*, *15*(S2), 93–101.

Mazor, K. M., Calvi, J., Cowan, R., Costanza, M. E., Han, P. K. J., Greene, S. M., et al. (2010). Media messages about cancer: What do people understand? *Journal of Health Communication*, *15*(S2), 126–145.

McCormack, L., Bann, C., Bann, L. S., Berkman, N., Squire, C., Schillinger, D. et al. (2010). Measuring health literacy: A pilot study of a new skills-based instrument. *Journal of Health Communication*, *15*(S2), 51–71.

Osborn, C. Y., Cavanaugh, K., Wallston, K. A., & Rothman, R. L. (2010). Self-efficacy links health literacy and numeracy to glycemic control. *Journal of Health Communication*, *15*(S2), 146–158.

Paasche-Orlow, M. K., & Wolf, M. S. (2010). Promoting health literacy research to reduce health disparities. *Journal of Health Communication*, *15*(S2), 34–41.

Parker, R., & Ratzan, S. (2010). Health literacy: A second decade of distinction for Americans. *Journal of Health Communication*, *15*(S2), 20–33.

Sarkar, U., Karter, A. J., Liu, J. Y., Adler, N. E., Nguyen, R., López, A., & Schillinger, D. (2010). The literacy divide: Health literacy and the use of an internet-based patient portal in the Diabetes Study of Northern California (DISTANCE). *Journal of Health Communication*, *15*(S2), 183–196.

Sudore, R. L., Schillinger, D., Knight, S. J., & Fried, T. R. (2010). Uncertainty about advance care planning treatment preferences among diverse older adults. *Journal of Health Communication*, *15*(S2), 159–171.

United Nations Economic and Social Council (ECOSOC). (2010). Health literacy and the Millennium Development Goals: UN Economic and Social Council (ECOSOC) Regional Meeting Background Paper (abstracted). *Journal of Health Communication*, *15*(S2), 211–223.

Wilson, E. A. H., Wolf, M. S., Curtis, L. M., Clayman, M. L., Cameron, K. A., et al. (2010). Literacy, cognitive ability, and the retention of health-related information about colorectal cancer screening. *Journal of Health Communication*, *15*(S2), 116–125.

Wynia, M. K., & Osborn, C. Y. (2010). Health literacy and communication quality in health care organizations. *Journal of Health Communication*, *15*(S2), 102–115.

Yost, K. J., Webster, K., Baker, D. W., Jacobs, E. A., Anderson, A., & Hahn, E. (2010). Acceptability of the talking touchscreen for health literacy assessment. *Journal of Health Communication*, *15*(S2), 80–92.

Journal of Health Communication, 15:34–41, 2010

ISSN: 1081-0730 print/1087-0415 online
DOI: 10.1080/10810730.2010.499994

Routledge
Taylor & Francis Group

Commentary

Promoting Health Literacy Research to Reduce Health Disparities

MICHAEL K. PAASCHE-ORLOW

Section of General Medicine, Department of Medicine, Boston University School of Medicine, Boston, Massachusetts, USA

MICHAEL S. WOLF

Northwestern University School of Medicine, Chicago, Illinois, USA

Limited health literacy has been linked to worse health outcomes for a range of medical conditions. In addition, limited health literacy is more prevalent among specific racial and ethnic minorities. Although these findings have been widely acknowledged, little systematic research has been conducted to elucidate the role of health literacy in the creation of health disparities or to evaluate the possibility that interventions relating to health literacy may help eliminate health disparities. This paper presents recommendations for a research agenda that is focused on advancing the science for how health literacy research can promote the effort to eliminate health disparities.

In the United States, racial/ethnic and socioeconomic disparities within the educational system have long been reported. As a result, two thirds of African American adults and 74% of Hispanic adults have limited functional literacy skills, compared to 32% of whites (Kutner, Greenberg, Jin, Boyle, Hsu, & Dunleavy, 2007). The consequences of early failures in education have more recently been linked to problems in healthcare. Research has begun to emerge showing how a health literacy skill set is linked to a range of health outcomes, and evidence has also emerged demonstrating how deficits in these skills possibly explain certain disparities (Bennett, 2009; Osborn, Cavanaugh, Wallston, White, & Rothman, 2009; Osborn, Paasche-Orlow, Davis, & Wolf, 2007; Sentell, 2006; Waldrop-Valverde, 2010; Wolf, 2006).

According to the most widely cited definition, health literacy is the "degree to which individuals have the capacity to obtain, process and understand basic health information and services needed to make appropriate health decisions" (Nielsen-Bohlman, 2004). As such, health literacy must be framed in the context of the specific tasks that need to be accomplished. In its seminal 2004 health literacy report *A Prescription to End Confusion*, the IOM recognized that patients' health-related knowledge, skills, and behaviors are heavily influenced by: (1) cultural

Address correspondence to Michael K. Paasche-Orlow, 801 Massachusetts Avenue, 2nd Floor, Boston, MA 02118, USA. E-mail: mpo@bu.edu

background, (2) health system demands, and (3) prior learning opportunities (Nielsen-Bohlman, 2004). The report frames limited health literacy not as a patient problem, but as a challenge to healthcare providers and health systems to reach out and more effectively communicate with patients. Specifically, the construct incorporates not only the individual cognitive and functional skills one uses when making health-related decisions, but it takes into account the contextual demands placed on the individual by (a) their specific clinical condition and associated health care decisions, (b) the communication characteristics of the dominant medical culture, (c) the structure and function of clinical services that assume adequate health literacy proficiency and require self-advocacy and vigilance, and (d) the emphasis that society places on individual, rather than ecological, determinants of health.

As one of the primary public health goals in the United States and other industrialized countries is to better understand and respond to health disparities, a health literacy perspective provides an important new direction for seeking perhaps more potentially modifiable strategies for reducing inequities in the short-term, with a multitude of targets. Health literacy researchers are now recognizing the need for comprehensive strategies that go beyond considering only a patient's functional literacy abilities. Rather, health literacy interventions should examine the complexity of the tasks required of patients and families within healthcare settings, the accessibility of providers for the target populations, the preparedness of health and public health professionals to engage productively with patients, and the features of the health care system, working environments, and communities in which care-giving and self-management support take place.

Perhaps the most significant attribute of health literacy research is that it calls attention to the many ways in which unnecessarily complex healthcare exacerbates the impact of underlying educational and income disparities. The implications of this extensive body of literature should be understood as a direct challenge to health systems that have been organized for the most highly educated and affluent members of a society. While seminal reports about the problem of limited health literacy have been issued by the Institute of Medicine (Nielsen-Bohlman, 2004), Agency for Healthcare Research and Quality (Berkman, 2004), American Medical Association (1999) and Joint Commission among others (Murphy-Knoll, 2007), none have addressed how health literacy research may help eliminate disparities.

Herein, we have sought to present recommendations for a research agenda that is focused on advancing the science for how health literacy research can promote the effort to eliminate health disparities and thereby promote health care equity. As this field is currently insufficiently developed, and there are many things that are not yet known about eliminating health disparities, the agenda presented here should be viewed as a work in progress.

1. Integrate Health Literacy Assessment in Disparities Research

The first step in most circumstances is to measure health literacy. Without measurement, it is not possible to know when and how health literacy may be relevant, and it would be very easy to design interventions that fail to attend to relevant factors. For example, in an adjusted analysis that excluded health literacy, African Americans were 2.4 times more likely to be non-adherent to their HIV medication regimen than whites (95% confidence interval [CI] = 1.14–5.08). When health literacy was included in the final model adjusting for relevant covariates, the effect estimates of race

diminished to non-significance and health literacy remained a significant independent predictor of non-adherence (adjusted odds ratio [AOR] = 2.12, 95% CI = 1.93–2.32) (Osborn, Paasche-Orlow, Davis, & Wolf, 2007). In another study, patients were asked their preferences regarding end-of-life care if they would develop advanced dementia. The African American subjects were found to prefer more aggressive care than the white subjects. When health literacy was included in the final model, health literacy—but not race—significantly predicted of preferences for care (low literacy OR 7.1, 95% CI 2.1–24.2) (Volandes et al., 2008). Health literacy clearly mediated the influence of race on end-of-life preferences. In both of these investigations, completely different conclusions would have been made without concurrent evaluation of race and health literacy.

2. Improve Patient Education

Limited health knowledge has been the most widely identified outcome in health literacy research, followed by certain health behaviors (Berkman, 2004). Given the strong associations between race, educational attainment, and health literacy, improving the timely access to information for specific health conditions or recommended actions among vulnerable populations is a logical step for promoting health equity (Ayotte, Allaire, & Bosworth, 2009; Bailey et al., 2009). Decades of research have produced an innumerable supply of publicly and commercially available patient education tools, available in an array of print, video, and web-based formats. Very few of these materials have incorporated a range of patient perspectives in their development, and even fewer have been formally tested to confirm their utility among culturally diverse and lower literate populations. While learning styles vary, efforts should be made to direct health providers and systems to the best available educational tools. This will likely require established criteria for evaluating their adequacy. To do this will be particularly challenging, as many tools that have already been developed specifically for lower literate audiences have not been entirely successful. Gerber et al. demonstrated a lack of significant knowledge gains for educating patients on diabetes self-care concepts using state-of-the-art multimedia (Gerber, 2005), and early efforts by Davis and colleagues found plain language print health information to produce only minimal improvement compared to existing brochures for polio vaccination (Davis, 1998). While guidance is available throughout various sources as to how best to create and field test plain language health information materials for diverse patient groups, there remains a need to consolidate recommendations, systematically assess the efficacy of newly available materials, and offer streamlined means to disseminate them at low or no cost to health care providers and systems. Equally important, finding ways to fund and produce language concordant versions of patient education tools for use among practices serving patients with limited English proficiency should be viewed as a public health priority.

3. Simplification of the Health Care System: Access and Utilization

Unneeded and complex barriers to access and utilize health services exacerbate the impact of underlying educational disparities. There is great potential for health insurance reform to improve the lives of millions of Americans. However, health insurance and other health benefit systems need to be designed to succeed.

Complex application forms, terms and conditions, and documentation requirements are significant barriers that disproportionately burden vulnerable populations (Davidoff, 2010; Ettner, 2010). A health literacy lens should be taken to all public programs: How will a person with limited literacy get and use this program? Answering such questions will help economically progressive programs fulfill their mission of reducing socioeconomic class disparities.

4. Simplification of the Health Care System: Education and Training of Health Professionals

Health professionals contribute to the unneeded complexity of the healthcare system with poor communication and limited dedication to patient education. Clinicians frequently use jargon (Castro, 2007) and rarely confirm if their patients understand what is being discussed (Schillinger, 2003). Consequently, patients frequently misunderstand a broad array of critical information (Fang, 2009). The Centers for Disease Control hosts a Web-based training program on health literacy that is geared toward public health (http://www.cdc.gov/healthmarketing/healthliteracy/training/) and the Health Resources Services Administration hosts a unified health communication course (http://www.hrsa.gov/healthliteracy/) to address Health Literacy, Cultural Competency, and Limited English Proficiency. These are free, on-line training programs to improve patient-provider communication. Integration of health literacy curricula into professional education, however, is in a nascent state. To date, there is a dearth of research on the impact of educational initiatives to promote the knowledge and skills of health professionals regarding health literacy. Curricular initiatives to address health disparities are further advanced. There is a National Consortium for Multicultural Education for Health Professionals including educators from 18 U.S. medical schools, which collect lessons learned from curriculum implementation to guide similar educational endeavors across the consortium (Carter-Pokras, 2010). Indeed, several states have laws and or regulations that mandate training health professions in cultural competence (https://www.thinkculturalhealth.org/cc_legislation.asp). Inclusion of health literacy in such policy initiatives can help promulgate health literacy curricula, but research will be needed to identify programs that work, can be replicated, and help eliminate health disparities.

5. Simplification of the Health Care System: Self-Care

Renewed interest in case management services, including care coordination, patient navigation, and certain proposed definitions of a medical home offer possible means for having lay workers or allied health professionals support patients in self care endeavors. Many longstanding, controversial health care practices that have previously been implicated as examples of poor health system design are now being confronted. For instance, solutions are being tested to manage the longstanding problem of variable and poor quality clinician prescribing practices that have been at the root of patient misunderstanding and prescription misuse. In fact, laws have now been set that lay out standards for both prescribing and labeling medicines to ensure patient-centered and evidence-based practices are in place to promote comprehension and adherence. In California and New York City, these regulations include providing language concordant spoken and text medication information (Bailey, Pandit, Curtis, & Wolf, 2009; "California State Board of Pharmacy,

Title 16-1707.5 Patient-Centered Labels on Medication Containers"). Indeed, there is a broad range of self-care tasks that could benefit from simplification. For example, ways to improve the FDA's Nutritional Facts label have been proposed, as Rothman and colleagues had previously reported more than half of adults can successfully navigate the current label and compute necessary information to support appropriate and safe decision making on food products (Lokker, 2009). There are many excellent targets for simplification and standardization as methods to improve self care and there remain great research opportunities in designing self-care systems that are effective for people with low health literacy. Moreover, very little is understood about the role of health literacy and its influence on health disparities for all the self-care tasks that occur outside of any relationship with a health provider. The public health arena has really been under explored.

6. Health Information Technology

There are increasing opportunities to leverage various health technologies to deliver health literacy-related interventions for a variety of purposes. Information technologies may range from automated telephone calls, to integrated electronic health record systems with patient portals and computerized agents. The value is clear; as health systems become automated, these tools allow for greater outreach, accountability, and standardization in the way health and healthcare is communicated to patients and families. Sarkar and colleagues compared the effectiveness of a diabetes self-management intervention using automated telephone calls with a nurse follow-up versus group medical visits and usual care (Sarkar, 2008). Those receiving the automated calls reported better perceived diabetes care, communication, and both intervention arms documented improved self-care behaviors, fewer bed days, and less interference with daily activities. In a small pilot of an animated, computerized agent to standardize communication for an informed consent process, Bickmore et al. (2009) found patients to be more accepting of the technology even compared to human interactions, although no differences in knowledge acquisition were noted. In the current issue Bickmore and colleagues extend this work to show that people with no prior computer experience and limited health literacy can readily use such a system. These studies highlight the potential value of health technologies for engaging patients, particularly those with limited health literacy; yet more research is needed to better document the link between these interventions and health outcomes. Greater attention to human-computer interaction considerations must also be made as Czaja (2008) aptly described the health literacy challenge to older adults in attempting to navigate and access information on the Medicare website.

7. Target the Educational System

For decades, studies have repeatedly highlighted the strong association between lower educational attainment and poor health outcomes, including mortality. These relationships have remained significant even after adjusting for relevant socioeconomic indicators. It would follow that as the quality of primary and secondary education improves and disparities in academic performance are reduced, so might health inequities. However, high school graduation rates have remained relatively stagnant in the United States for decades. Much remains to be learned about the pathways from education to health. Elucidating how general literacy, which can

influence among other things factors like employment options, financial stability, and access to health care, operates in similar or different ways in people's lives as health literacy could help direct efforts in designing education programs that improve health and decrease health disparities. Disparities in basic literacy skills may have a very different—and even bigger—role in promoting health disparities in the United States than disparities in health literacy skills. Improving the high school graduation rate and improving the quality of primary and secondary education should be seen as long-term public health measures. Another long-term goal should be to develop successful initiatives that advance the health literacy of future generations by increasing patient skills and public health skills across the lifespan, throughout all levels of the education system. School curricula should include common and pertinent medical terminology, practical healthcare navigation skills, and age-related health issues such as physical activity and nutrition. This can lead to more accurate expectations of current and future roles and responsibilities regarding self-care and prevention, as well as instilling the type of self-efficacy that may be needed to engage health problems as they arise.

Conclusion

The fact that low health literacy is associated with worse health outcomes and with specific racial and ethnic groups makes this field of inquiry particularly appealing as a source of ideas, methods, and interventions to help eliminate disparities. And yet, to date, the fields of health literacy research and health disparities research have remained largely separate. We have presented seven high priority areas for what we hope may amount to the beginning of an active shared health literacy and health disparities research agenda. The agenda we present needs to be interpreted contextually and may need to be recast to meet the needs of various public health and international settings.

References

Ayotte, B. J., Allaire, J. C., & Bosworth, H. (2009). The associations of patient demographic characteristics and health information recall: The mediating role of health literacy. *Neuropsychology, Development, and Cognition. Section B, Aging, Neuropsychology, and Cognition*, *16*(4), 419–432.

Bailey, S. C., Pandit, A. U., Curtis, L., & Wolf, M. S. (2009). Availability of Spanish prescription labels: A multi-state pharmacy survey. *Medical Care*, *47*(6), 707–710.

Bailey, S. C., Pandit, A. U., Yin, S., Federman, A., Davis, T. C., Parker, R. M., et al. (2009). Predictors of misunderstanding pediatric liquid medication instructions. *Family Medicine*, *41*(10), 715–721.

Bennett, I. M., Chen, J., Soroui, J. S., & White, S. (2009). The contribution of health literacy to disparities in self-rated health status and preventive health behaviors in older adults. *Ann Family Medicine*, *7*(3), 204–211.

Berkman, N. D., DeWalt, D. A., Pignone, M. P., Sheridan, S. L., Lohr, K. N., Lux, L., Sutton, S. F., et al. (2004). *Literacy and health outcomes. Evidence report/technology assessment No. 87 (AHRQ Publication No. 04-E007-2)*. Rockville, MD: Agency for Healthcare Research and Quality.

Bickmore, T. W., Pfeifer, L. M., & Paasche-Orlow, M. K. (2009). Using computer agents to explain medical documents to patients with low health literacy. *Patient Education and Counseling*, *75*(3), 315–320.

California State Board of Pharmacy, Title 16-1707.5 Patient-Centered Labels on Medication Containers. Retrieved January 25, 2010, from http://www.pharmacy.ca.gov/laws_regs/1707_5_proposed_text.pdf

Carter-Pokras, O., Bereknyei, S., Lie, D., & Braddock, C. H. 3rd; National Consortium for Multicultural Education for Health Professionals. (2010). Surmounting the unique challenges in health disparities education: A multi-institution qualitative study. *Journal of General Internal Medicine*, *25*(Suppl 2), S108–S114.

Castro, C. M., Wang, F., & Schillinger, D. (2007). Babel babble: Physicians' use of unclarified medical jargon with patients. *American Journal of Health Behavior*, *31*(Suppl 1), S85–S95.

Czaja, S. J., & Nair, S. N. (2008). Usability of the Medicare health web site. *Journal of the American Medical Association*, *300*(7), 790–792.

Davidoff, A. J., Shaffer, T., Shoemaker, J. S., Kim, M., & Zacker, C. (2010). Lessons Learned: Who Didn't Enroll in Medicare Drug Coverage in 2006, And Why? *Health Affairs (Millwood)*, *29*(6), 1255–1263.

Davis, T. C., Arnold, C., Murphy, P. W., Herbst, M., & Bocchini, J. A. (1998). A polio immunization pamphlet with increased appeal and simplified language does not improve comprehension to an acceptable level. *Patient Education and Counseling*, *33*(1), 25–37.

Ettner, S. L., Duru, O. K., Turk, N., Quiter, E., Schmittdiel, J., & Mangione, C. M. (2010). Entering and exiting the medicare part d coverage gap: Role of comorbidities and demographics. *Journal of General Internal Medicine*, *25*(6), 568–574.

Fang, M. C., Machtinger, E. L., & Schillinger, D. (2009). Language, literacy, and characterization of stroke among patients taking warfarin for stroke prevention: Implications for health communication. *Patient Education and Counseling*, *75*(3), 403–410.

Gerber, B. S., Brodsky, I. G., Lawless, K. A., Smolin, L. I., Arozullah, A. M., Smith, E. V., et al. (2005). Implementation and evaluation of a low-literacy diabetes education computer multimedia application. *Diabetes Care*, *28*(7), 1574–1580.

Health literacy: Report of the Council on Scientific Affairs. Ad Hoc Committee on Health Literacy for the Council on Scientific Affairs, American Medical Association. (1999). *Journal of the American Medical Association*, *281*(6), 552–557.

Kutner, M., Greenberg, E., Jin, Y., & Paulsen, C. (2006). *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.

Kutner, M., Greenberg, E., Jin, Y., Boyle, B., Hsu, Y., & Dunleavy, E. (2007). *Literacy in everyday life: Results from the 2003 National Assessment of Adult Literacy (NCES 2007–480). U.S. Department of Education*. Washington, DC: National Center for Education Statistics.

Lokker, N., Perrin, E. M., Kumar, D., Finkle, J., Franco, V., Choi, L., Johnston, P. E., & Rothman, R. L. (2009). Parental misinterpretations of over-the-counter pediatric cough and cold medication labels. *Pediatrics*, *123*(6), 1464–1471.

Murphy-Knoll, L. (2007). Low health literacy puts patients at risk: The Joint Commission proposes solutions to national problem. *Journal of Nursing Care Quality*, *22*(3), 205–209.

Nielsen-Bohlman, L., Panzer, Allison M., & Kindig, David A. (2004). *Health literacy: A prescription to end confusion*. Washington, DC: Institute of Medicine: National Academies Press.

Osborn, C., Cavanaugh, K., Wallston, K. A., White, R. O., & Rothman, R. L. (2009). Diabetes numeracy: An overlooked factor in understanding racial disparities in glycemic control. *Diabetes Care*, *32*(9), 1614–1619.

Osborn, C. Y., Paasche-Orlow, M. K., Davis, T. C., & Wolf, M. S. (2007). Health literacy: An overlooked factor in understanding HIV health disparities. *American Journal of Preventive Medicine*, *33*(5), 374–378.

Sarkar, U., Gupta, R., Tang, A., Murphy, E., Seligman, H. K., Shojania, K. G., & Schillinger, D. (2008). Use of an interactive, telephone-based self-management support program to

identify adverse events among ambulatory diabetes patients. *Journal of General Internal Medicine*, *23*(4), 459–465.

Schillinger, D., Grumbach, K., Wang, F., Wilson, C., Daher, C., Leong-Grotz, K., Castro, C., & Bindman, A. B. (2003). Closing the loop: Physician communication with diabetic patients who have low health literacy. *Archives of Internal Medicine*, *163*(1), 83–90.

Sentell, T., & Halpin, H. A. (2006). Importance of adult literacy in understanding health disparities. *Journal of General Internal Medicine*, *21*(8), 862–866.

Volandes, A. E., Paasche-Orlow, M., Gillick, M. R., Cook, E. F., Shaykevich, S., Abbo, E. D., et al. (2008). Health literacy not race predicts end-of-life care preferences. *Journal of Palliative Medicine*, *11*(5), 754–762.

Waldrop-Valverde, D., Osborn, C. Y., Rodriguez, A., Rothman, R. L., Kumar, M., & Jones, D. L. (2010). Numeracy skills explain racial differences in HIV medication management. *AIDS Behavior*, *14*(4), 799–806.

Wolf, M. S., Lyons, E. A., Durazo-Arvizu, R., Pickard, S. A., Arseven, A., Arozullah, A., Colella, K., Ray, P., & Bennett, C. L. (2006). Literacy, race, and PSA level among low-income men newly diagnosed with prostate cancer. *Urology*, *68*(1), 89–93.

HSR Health Services Research

DOI: 10.1111/j.1475-6773.2009.01079.x
METHODS ARTICLE

Patient Navigation: Development of a Protocol for Describing What Navigators Do

Victoria A. Parker, Jack A. Clark, Jenniffer Leyson, Elizabeth Calhoun, Jennifer K. Carroll, Karen M. Freund, and Tracy A. Battaglia

Objective. To develop a structured protocol for observing patient navigators at work, describing and characterizing specific activities related to their goals.
Data Sources/Setting. Fourteen extended observations of navigators at three programs within a national trial of patient navigation.
Study Design. Preliminary observations were guided by a conceptual model derived from the literature and expert consensus, then coded to develop and refine observation categories. These findings were then used to develop the protocol.
Methods. Observation fieldnotes were coded, using both a priori codes and new codes based on emergent themes. Using these codes, the team refined the model and constructed an observation tool that enables consistent categorization of the observed range of navigator actions.
Findings. Navigator actions across a wide variety of settings can be categorized in a matrix with two dimensions. One dimension categorizes the individuals and organizational entities with whom the navigator interacts; the other characterizes the types of tasks carried out by the navigators in support of their patients.
Conclusions. Use of this protocol will enable researchers to systematically characterize and compare navigator activities within and across programs.

Key Words. Continuity of patient care, social support, health services needs and demands, case management, patient navigation

Challenges arise when people contemplate cancer screening. They multiply when tests suggest a threatening disease and indicate the need for follow-up investigations. And the challenges expand dramatically when such investigations reveal cancer, as people become "patients" in the complex realm of referrals, consultants, examinations, decisions, and often arduous treatment regimens.

Individuals who are socially and economically disadvantaged may find cancer care all the more problematic (Baquet et al. 2005). These people are at substantial risk of receiving inadequate care at each step of the cancer care continuum: screening, diagnostic follow-up of suspicious results, treatment when cancer is diagnosed, and survivorship surveillance. Moreover, the systems of care available to them, such as "safety net" institutions, are often beset with inefficiencies.

Patient navigation has emerged in the past decade in response to these widely recognized disparities in cancer care. Health care advocates, policy makers, and innovative health care organizations have called for the adoption of patient navigation to assist patients and remedy inefficiencies in the provision of timely care (Freeman 2006; Jandorf et al. 2006). As a result, patient navigation services have proliferated rapidly in recent years.

Research to date supports the promise of patient navigation for reducing cancer disparities (Gabram et al. 2008). Currently, a large, multisite cooperative study of patient navigation is being conducted. The National Cancer Institute's Patient Navigation Research Program (PNRP) is designed to evaluate the effectiveness of navigation in improving timeliness of care (i.e., time to follow-up of abnormal screening results and to completion of treatment when cancer is diagnosed) and patient satisfaction (Freund et al. 2008). The nine cooperating studies of the PNRP provide a laboratory for characterizing the actual work of patient navigators and linking variation in their activities to patient outcomes. The first step in such an effort is to design

---

Address correspondence to Victoria A. Parker, D.B.A., Assistant Professor, Department of Health Policy and Management, Boston University School of Public Health, 715 Albany St. T3W, Boston, MA 02118-2526, e-mail: vaparker@bu.edu. Victoria A. Parker, D.B.A., Assistant Professor, is with the Center for Organization, Leadership and Management Research, VA Boston Healthcare System, Boston, MA. Jack A. Clark, Ph.D., Associate Professor, is with the Center for Health Quality, Outcomes, and Economic Research—VA HSR&D, Edith Nourse Rogers Memorial Veterans Hospital, Bedford, MA and the Department of Health Policy and Management, Boston University School of Public Health, Boston, MA. Jenniffer Leyson, M.A., Senior Qualitative Data Analyst, is with the Department of Health Policy and Management, Boston University School of Public Health, Boston, MA. Elizabeth Calhoun, Ph.D., Associate Professor and Senior Research Scientist, is with the Division of Health Policy and Administration, University of Illinois School of Public Health, Chicago, IL. Jennifer K. Carroll, M.D., M.P.H., Assistant Professor, is with the Department of Family Medicine and Oncology, University of Rochester Medical Center, Family Medicine Research Programs, Rochester, NY. Karen M. Freund, M.D., M.P.H., Professor of Medicine and Epidemiology, and Tracy A. Battaglia, M.D., M.P.H., Assistant Professor of Medicine and Epidemiology, are with 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.

procedures for systematically observing navigators' activities. This paper reports on the development of a protocol for observing what navigators actually do.

BACKGROUND

Despite continued advances across the spectrum of cancer care (Brenner, Gondos, and Arndt 2007), the distribution of these advances remains uneven. They are less likely to be enjoyed by those segments of our society defined by minority racial and ethnic status, low income, and limited health insurance (Weir et al. 2003; Shavers, Fagan, and McDonald 2007). Disparities in cancer care are persistent and may in some instances actually be widening (Ries et al. 2007). Inequitable outcomes may result from, among other factors, well-documented delays in accessing diagnostic and treatment services by the most at-risk populations (Chang et al. 1996; Peterson, Han, and Freund 2003; Battaglia et al. 2007).

Several seminal reports (Smedley et al. 2003; Weir et al. 2003) have highlighted the barriers to cancer care inherent to socioeconomic disadvantage. Patient navigation is a community-based approach to reducing these barriers (Dohan and Schrag 2005; Battaglia et al. 2007; Ell et al. 2007; Ferrante, Chen, and Kim 2008). Guided both by principles of disease management and by cultural sensitivity, navigators are responsible for identifying individuals most at risk for delays in cancer care and mitigating barriers to their receipt of that care (Vargas et al. 2008). Navigation programs also seek to remedy systemic barriers to care within organizations delivering care. Patient navigation services address barriers by assigning trained supportive staff who track patients and assist them in completing their diagnostic and treatment care, while also advocating for solutions to systemic causes of those barriers.

Navigation programs are usually funded through local resources or foundation support because insurers do not reimburse this care. Local innovation results in tremendous variability in program structures and activities. Trailing behind these developments is a small, but growing body of research documenting the efficacy of navigation (Battaglia et al. 2007; Ell et al. 2007; Ferrante, Chen, and Kim 2008; Gabram et al. 2008). While these studies provide evidence that navigation is effective, the key components of a successful navigation program are not well understood.

Conceptual Definition

Problems often arise in the evaluation of complex innovations such as patient navigation because the interventions have not been fully defined and developed (Campbell et al. 2000). Early evaluations of innovative programs often simply assume either that the intervention is in place as planned or that some variation is acceptable (Eccles et al. 2003). However, such assumptions lead evaluators to overlook the effects of variation. And in the case of patient navigation, a lack of information about variation in definition, style, and scope could lead to inaccurate conclusions about its effectiveness.

There is no generally accepted definition of patient navigation. Reviewing 56 articles published before early 2004, Dohan and Schrag (2005) identify two types of definitions: "service focused" and "barrier focused." Service-focused definitions attend to activities such as connecting individuals to resources and assisting patients in completing courses of care. Dohan and Schrag (2005) criticize these definitions as nonspecific: such activities could be, and often were, performed by other providers as part of their duties. Barrier-focused definitions, they argue, attend to activities that identify and remove impediments preventing patients from moving through screening, diagnostic follow-up, and treatment. Furthermore, these responsibilities were distinguishable from those usually assigned to social workers, case managers, community outreach workers, and health advocates. The latter roles, they argue, were typically proactive, providing education and counseling services, while navigation was essentially reactive to emergent impediments to care.

However, defining patient navigation in terms of resolving barriers for individual patients may be too constricting. While service-oriented definitions could blur distinctions between navigators and other service providers, focusing only on what navigators do in relation to barriers facing individual patients may obscure a variety of related activities they perform. They may tweak organizational practices to expedite patient care, develop local resources for multiple patients, and build cooperative relationships with and among clinic staff that facilitate more efficient movement of patients through systems. Thus, while staff other than navigators may facilitate patient access to services, to define patient navigation work solely in terms of barrier reduction risks artificially excluding other patient navigation functions.

We sought to avoid such exclusions. Informed by research to date, we noted that navigators work with health care organizations—sometimes within, sometimes externally—to facilitate patients' receipt of care from providers. This framing gives attention to the fact that navigators often must involve

others in their work, which, in turn, suggests that navigators' networks of relationships with these "others" might be essential to achieving their objectives. Drawing on research in various care settings, we then defined navigation in terms of tasks and networks: *navigators do things for patients by working with patients and other actors in both the social network of the organization itself and the community in which the organization resides.*

Social networks have been described as "patterns of relations joining actors" (Marsden 1990). Keating et al. (2007) use the social network approach to understand patterns of advice-giving and -following within a primary care practice. Earlier navigation research emphasized the pivotal role of the navigator in helping patients access necessary services, which suggests that part of navigation is knowing to whom to go for specific support. Thus, to understand what navigators do, we must understand their patterns of relations with others who provide services that facilitate screening, diagnosis, and/or treatment. Underscoring the importance of understanding these networks is a recent analysis of early patient navigation programs that defines patient navigation as a system, rather than a person (Vargas et al. 2008). The network framework is useful in bridging the service-focused/barrier-focused dichotomy: it usefully attends to the question of *how* navigators achieve their efforts on behalf of patients. Whether obtaining services proactively or in response to a specific barrier, navigators engage others in their networks to find, arrange, and seek reimbursement for those services.

The social network concept illuminates part, but not all, of the scope of patient navigation. We need also to describe the activities of navigators. Task analysis, with its emphasis on interaction of persons and environment, offers a useful complement. Task analysis is concerned with identifying the goal of a task, the criteria for reaching that goal, and the relevant resources and constraints. It also emphasizes that task-directed actions are determined by both the person carrying them out and the relevant environment (which presumably includes the person's social network), and that people develop personal expertise in how to accomplish their tasks (Norros and Nuutinen 2002). Navigators need to build a working knowledge of the tasks they must perform and a network of contacts to support their actions.

METHODS

Development of the observation protocol was grounded in a qualitative study of PNRP navigators at work. We began by surveying the nine PNRP sites to

characterize structural attributes of each program that would define the contexts in which navigators worked. Attributes included each program's physical site (facilities, geographic location, and populations served), its size, and the spectrum of navigation services offered. Based on these findings, we selected a convenience sample of three pilot programs that represented the diversity of settings and approaches to navigation implemented at the nine independent sites.

We also developed a preliminary guide for observing navigators at work. This guide was designed to enable the collection of comparable field observations of what navigators do (tasks) and the people and entities with whom they interact (networks) in accomplishing those tasks. The resulting data were analyzed to develop a comprehensive, yet simple observation protocol for observation of navigators across all PNRP sites and elsewhere.

Sample

The nine programs in the PNRP were selected through a competitive national process and designed according to general program criteria set by the NCI (Freund et al. 2008). The NCI sought to maximize diversity within this group to assess the usefulness of patient navigation across a range of settings. Thus, the programs differed in many respects, but they met common requirements regarding navigator training, patient population, patient criteria for inclusion, and collection of data. The programs address different combinations of four cancers—breast, cervical, prostate, and colorectal—where navigation would likely have a detectable effect in facilitating follow-up of suspicious screening results and completion of treatment (Freund et al. 2008). Each of these cancers is associated with a distinct patient population and pattern of care.

The locations where navigators work and interact with patients, medical providers, and others vary: hospital evaluation clinics, inpatient wards, and treatment units; primary care clinics; and community health centers. The scope of navigator involvement relative to the cancer care continuum also varies; some navigate in all phases, while others navigate only from diagnosis through treatment. At some sites, navigators focus on case finding, while at others, the focus is on supporting patients through treatment. The number of navigators employed varies, as does navigators' involvement in competing clinical or administrative responsibilities. Some navigators were hired and are supervised directly by the research program; others were hired by the clinical care sites themselves as a subcontract to the research program. Finally, the programs vary with respect to the professional background and training of

navigators: some use navigators with clinical training and credentials, while others use "lay" navigators selected for congruence with the target patient population.

In selecting the pilot sites, we excluded one program due to restrictive local access requirements. Site 2 was selected because it is the research team's home site and had a small, longstanding navigation program that predates the PNRP. Sites 4 and 7 were selected to provide informative, qualitative contrasts along the dimensions outlined above. Information about all eight programs is presented in Table 1.

Preliminary Data Collection

We enlisted other PNRP investigators to collaborate at their respective sites. Multisite, multi-investigator qualitative research required developing a common protocol. We designed a semistructured observation guide to support collection of comparable data across all sites. During a 1-day training, observers from each site were directed to describe specific actions of navigators and note the following: (1) the approximate duration of actions; (2) the parties with whom navigators interacted; (3) whether interactions were in person or via phone/email; (4) the relevance of the action to navigation.

Observers also were directed to query navigators about their actions at moments that would not interrupt observed activities. Navigators were asked to explain the relevance of observed activities to particular navigation issues and challenges, including (1) actions navigators took to develop a relationship with the patient and others relevant to the patient's case; (2) the role of others who were consulted for advice, direction, or assistance; (3) the initiation and extent of an interaction with a person; (4) the nature of the problems being addressed. Thus, while we focused on directly observable behaviors, we also explored navigators' reflections on the scope of and rationale for their actions.

Each navigator at each site was observed at least twice. Observations were scheduled in consultation with navigators to best capture the variation in their workflow, so the lengths of the sessions were defined by the ways navigators scheduled their own activities. For example, if a navigator worked with a specific provider seeing patients for screening follow-up, the observation ran the length of the provider's clinic. Most observations lasted about 4 hours.

At the end of each observation, navigators were asked further questions to characterize the representativeness of the actions just observed:

- How did you decide what to spend time on today?

Table 1: Characteristics of Patient Navigation Research Program (PNRP) (Protocol Development Sites Are Highlighted)

Program Location	*Cancer Targets*	*Number of Sites Where Navigation Is Offered*	*Type of Site(s) Where Navigator Is Based*	*Number of Navigators*	*Minimum Navigator Education Requirement (Set at Site)*	*Navigators Have Other Clinical/Administrative Responsibilities*
1	Breast	6	Hospital-based patients, hospital outpatient clinics, community health centers (CHCs)	8	None	Yes
2	Cervical, breast	6	CHCs	6	High school diploma	Yes
3	Breast, colorectal	10–12	CHCs	3	High school diploma	No
4	Cervical, breast, prostate, colorectal	5	CHCs, VA outpatient clinic	5	High school diploma	Yes
5	Breast, cervical	2	Hospital outpatient clinics, CHCs	4	BS/BA or LPN	No
6	Breast, colorectal, prostate	2	Specialty clinics at hospital	5	High school diploma, 2 years of work experience in medical setting	No
7	Breast, colorectal	11	CHCs, hospital outpatient oncology clinics and treatment units, primary care clinics	3	Completion of local Family Development Credential Program and in-house training	No
8	Breast, cervical, colorectal	6	Hospital outpatient clinic, CHCs	3	BA/BS and relevant work experience	No
9	Breast, cervical, prostate, colorectal	4	CHCs	4	PNRP and local training	Yes

- How did this action relate to prior activities for a particular patient or patients in general?
- Does this action reflect any specific strategies for assessing/obtaining what the patient needs?
- What will you do next?
- Did the problematic situation/barrier get resolved?
- How typical a day was this for you?

The observation guide was refined in a 1-day meeting of the observers, followed by regular conference calls. Issues of reliability were addressed by comparing observations within and between the three sites. Apparent discrepancies in the quality of observations were discussed and resolved by developing a consensus among all observers.

Analysis

Using the observation guide, investigators conducted a total of 18 comparable observations of nine navigators working in three program sites of the PNRP. The analysis focused on refining the observation protocol. Data from all three sites were compiled by the project PIs. Fieldnotes from each observation were imported into software supporting text-based analysis, HyperRESEARCH (ResearchWare Inc. 2008).

The analysis was informed by the general approach of grounded theory methodology (Glaser and Strauss 1967). The first three authors each reviewed separately one set of fieldnotes from each of the three sites, identifying themes that characterized the activities reported. The team then met to compare descriptive codes and reach consensus on code definitions before coding the rest of the fieldnotes. The primary aim of this analysis was to define a comprehensive, yet parsimonious set of categories that would enable other observers to reliably categorize navigator behavior.

Once these categories were developed, we also drafted observation instructions to guide subsequent field observations of the work of navigators. Drafts of the analytic categories were developed by the Site 2 researchers and shared with the Sites 4 and 7 observers for substantive critique and revision in a day-long research team meeting, which yielded the final protocol reported herein.

RESULTS

The three sites provided a wide array of contexts for observing navigation. They differed with respect to the scope of the navigation program, the phases of cancer care addressed (i.e., screening, diagnosis, or treatment), the history/longevity of the program, the emphasis placed on various navigator responsibilities, and the background (e.g., clinical, survivor, cultural/ethnic) of the navigators, as well as their physical and organizational location (e.g., community health centers, large medical centers, outpatient primary care, or treatment clinics). These contextual differences appeared to influence, to an undetermined extent, what navigators do. Thus, we sought an observation protocol that would reliably capture activities in these diverse settings.

Domains of Navigator Behavior

Guided by the concepts of *task* and *network*, we defined five categories each of navigator tasks and social networks. The task categories include navigating with a patient, facilitating for a patient, maintaining systems for all patients, documenting/reviewing actions, and other tasks. The five network categories include patient(s), clinical provider(s), nonclinical staff, formal and informal support, and medical record systems. Each of these categories is defined, described, and illustrated below.

Task Categories. Navigating tasks consist of identifying and mitigating barriers *with* patients. They include telling (explaining when and where biopsy will be done, describing what it will be like); inquiring (asking about barriers to attending the appointment, exploring the patient's concerns); supporting (listening to fears about treatment); and coaching (discussing questions that need to be asked at next appointment and how to ask them).

Facilitating tasks are performed *for* a specific patient. They include finding (locating current patients and ensuring that they will come to appointments); coordinating team communication (ensuring the entire care team is aware of the next steps); integrating information (ensuring that different types of patient data are documented and shared as needed); and seeking collaboration (enlisting other providers in addressing the patient's fears).

Maintaining systems tasks support *all* patients. They include identifying potential patients (reviewing lab results to note patients who need follow-up); building networks and referral routines (meeting with clinicians to explain

navigator role and discuss referral criteria); and reviewing cases (checking on ticklers and open issues).

Documenting activities and *reviewing* information constitute another major navigator task. They include recording navigator actions (recording steps taken with or on behalf of the patient in the patient's medical record or a separate navigation file); handling test results (retrieving and entering patient data from labs, radiology, or other sources); and processing other necessary information (recording information or activities relevant to navigator role).

Other activities are those apparently unrelated to navigation. It was important to capture all network interactions, even when their relevance to navigation was not apparent. For example, many navigators have other distinct roles unrelated to navigation; documenting these other activities will help in understanding how the navigator role fits in with other roles, both formally and informally. This category includes research-related activities, such as consenting patients, providing clinical back-up, activities unrelated to navigation (interpreting for nonnavigated patients), and socializing (having informal conversation with co-workers).

Network Categories. Navigators may interact with a specific *patient*, such as when phoning the patient with information about an upcoming diagnostic procedure.

Navigators may also interact with *providers*, both within and outside their immediate location. For example, s/he might speak with the physician to confirm the meaning of a test result before discussing it with the patient.

Nonclinical staff, such as receptionists or administrators coordinating insurance, represent another group with whom the navigator may interact.

People who provide *supportive services*, either formally (social workers, translators, transportation staff) or informally (friends, family) within or outside the facility are another group with whom navigators interact.

The final category—*paper or electronic medical record systems*—could be perceived as merely a means to communicate with members of the other four network categories, and it does function in that way. However, our preliminary observations indicated that, in the eyes of the navigator, the medical record itself takes on some of the qualities of a person, in that it needs to be informed and/or consulted before other actions are taken. This observation is consistent with those of many studies of human–computer interaction (Turkle 2003).

Observation Protocol Refinement

The current observation protocol incorporates solutions to problems encountered in the field using the preliminary observation guide. Initially, observers were required to take continuous notes, recording the duration and mode of the navigator's activity, the person with whom s/he spoke, the activity, and the patient on whose behalf the activity was taken, plus descriptive narrative. This recording burden proved too onerous in the field: recording all observed activities not only interfered with the primary goal of noting tasks and social networks used by the navigators, but it did not produce more useful data.

Based on this early finding, two important changes were made to the observation protocol: activities were observed in 15-minute intervals, and coding focused on the primary activity of each interval. This time sampling methodology facilitates detailed reporting of navigator activities without attempting to capture everything that occurs during an observation. Observers start a new form every 15 minutes, focusing notes and coding on the navigator's primary activity during that period. Thus, each hour of observation time yields four distinct chunks of description and activity coding. This sampling interval provides some sense of the relative proportion of a navigator's time spent on different activities, while allowing observers to record more detailed notes about the main activity. This approach necessarily involves some observer judgment: sometimes a navigator tackles multiple short tasks during a single 15-minute interval. In such instances, observers were instructed to either group-like tasks into a single entry (making appointment reminder calls to a list of patients could be meaningfully described as one task) or focus on the first activity during the time period.

Through discussion, a five-by-five matrix emerged, with tasks on the vertical axis and social networks on the horizontal axis. The observation form itself was redesigned to incorporate on a single page both this simple matrix and an open area for handwritten fieldnotes (see Figure 1).[1]

After field testing, several additional refinements were made. Certain combinations of tasks and networks cannot occur. For example, the task of navigating can be performed only with a patient, while reviewing a patient's file can be done only with the medical record. To further simplify the form, matrix cells representing combinations that cannot occur are blacked out.

Also, observers at some sites reported that a significant amount of navigation is carried out by telephone, leaving and returning voicemail messages. Therefore, for each observation of a navigator action that involves contact with a patient, the observer also notes whether the interaction is *synchronous*

Figure 1: Protocol Data Collection Form

*** Begin a new form every 15 minutes ***

Fieldnotes:

For each task observed, mark the appropriate empty cell. (Darkened cells are not in use.)

		1	2	3	4	5
	Tasks/Network	**Patient (S or A)**	**Provider**	**Non-clinical staff**	**Supportive services**	**Med rec / EMR**
A	Navigate with specific patient (tell, inquire, support, coach)					
B	Facilitate for specific patient (bring pt in, coordinate comm, seek advice)					
C	Maintain system for all patients (ID potential pts, build int/ext networks)					
D	Document/Review (record info, actions, results)					
E	Other (do research, provide clinic back-up, do non-nav tasks, socialize)					

If you cannot decide what box to mark, check here: _____

(happening in real time) or *asynchronous* (delayed, as when leaving a voicemail) by recording either "S" or "A" in the appropriate cell. For all other cells in the matrix, a checkmark is used.

Finally, because some observed activities may involve more than one person or task, observers are encouraged to mark more than one cell if that

best reflects what they are seeing. For example, if the navigator accompanies a patient to a physician visit, the observer puts an "S" in the cell representing "navigate/patient" and a checkmark in the box representing "facilitate/provider."

The matrix supports coding of real-time activities as they occur, but we realized the need for simultaneous, structured, narrative fieldnotes, as well. Hence, we developed observation guidelines directing the observer to note relevant contextual factors, such as the location of navigation activity, the language used in navigation, the racial/ethnic backgrounds of both patient and navigator if known, and the navigator's other roles (if any) in the organization.

This matrix facilitates rapid categorization of tasks and networks, allowing the observer to concentrate on writing narrative description that will document important information about context and content that cannot be fully captured by the matrix. The observer is encouraged to ask the navigator questions to develop a better understanding of what the navigator is doing and why.[2] Observers also are asked to record their impressions about interactions, clearly identifying these notes as their perceptions. For example, the observer might write "navigator and patient embraced warmly and seem to know each other well." While these impressions are particular to specific observers, they nevertheless add richness to the description.

DISCUSSION

Patient navigation represents an emerging innovation in care adapted to a variety of specific local contexts. Capturing local adaptations is crucial to meaningfully assess the efficacy of navigation across different sites. The protocol we have developed reflects a plausible, generic definition of navigation that has been found thus far to be applicable in multiple contexts. We expect the protocol to be useful in capturing the existing variation in navigation programs, and we will use it for this purpose as our research program goes forward. This type of data is essential to inform both research and practice.

While this protocol represents an important step, it certainly does not capture every detail of navigators' actions. For example, while interactions with other providers are categorized and noted, the protocol does not note how extensive or collaborative they might appear to be. Data from narrative fieldnotes will compensate for this limitation, while also providing the potential for further revisions to the protocol based on emerging patterns in these data.

This protocol is grounded in a qualitative study of navigation in three sites of the PNRP. As such, it has a measure of validity, yet its validity requires further examination through application. The current definitions of the task/network categories may reflect the particular realities of the sites we have observed in developing the protocol; they may evolve as new sites are studied. Likewise, variation in organizational, political, and community contexts in which operational navigation programs are developed may require modifications to this protocol. Moreover, while this protocol enables the systematic observation of what navigators do, it is only part of a comprehensive method for evaluating the processes and outcomes of patient navigation. The mix of patients served, their resources and ability to access care; the types of health problems for which access is needed; and the specific array of health services and providers for which navigation is needed must be taken into account in evaluating the effectiveness of what navigators do, as captured by this, or any protocol.

To further investigate the protocol's validity, we are implementing it on a wider scale. It is being used at eight sites to produce a dataset of approximately 130 observations (four half-day observations of each navigator at each site). As of this writing, 89 observations have been completed, and none has presented activities that fall outside the task/network categories described above. Quantitative and qualitative analyses of the data collected will enable the research team to characterize variation, both within and across sites, in navigator tasks, networks, and emphasis.

Protocol development thus far has illuminated important dimensions along which navigation programs may vary. By accurately characterizing this variation, researchers should be better able to interpret variation in patient outcomes associated with different navigation programs. While this protocol does not provide information on program effectiveness, it provides important process information that may help explain the connections between navigation context and outcomes. Of course, outcomes also may be affected by the actions of other health care providers and advocates, which may overlap with the actions of navigators. It is important not to over-attribute outcomes to the actions of specific navigators, but rather to keep analysis at the programmatic level. In other words, while patient navigator actions may often be directed at specific patients, navigation is properly conceptualized as a systemic intervention that changes how care is delivered. Thus, any change in outcomes observed may be the result of the actions of multiple individuals, including the navigator, whose actions have been influenced by the presence of the navigation program. The information obtained from further use of this protocol

to study the work of navigators may also inform the processes of selecting, training, supervising, and supporting navigators, as it will illuminate different practices that might optimize desired navigation program outcomes.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: The research reported in this paper was supported by a grant from the Avon Foundation to Boston Medical Center (BMC), which subcontracted with the research teams at University of Chicago (Chicago) and University of Rochester Medical Center (Rochester). Battaglia is principal investigator of the project. Parker is co-principal investigator; Clark, co-investigator; Freund, clinical consultant; and Leyson, research assistant. All authors except Parker are also principal investigators, co-principal investigators, or research staff at their respective participating sites in the National Cancer Institute's Patient Navigation Research Project (PNRP).

Disclosures: None.

Disclaimers: None.

NOTES

1. The observation form and matrix are reproduced here with the permission of the Trustees of Boston University.
2. Please contact authors for observer training instructions.

REFERENCES

Baquet, C. R., K. M. Mack, J. Bramble, M. DeShields, D. Datcher, M. Savoy, K. Hummel, S. I. Mishra, S. E. Brooks, and S. Boykin-Brown. 2005. "Maryland's Special Populations Cancer Network: Cancer Health Disparities Reduction Model." *Journal of Health Care for the Poor and Underserved* 16: 192–206.

Battaglia, T. A., K. Roloff, M. A. Posner, and K. M. Freund. 2007. "Improving Follow-Up to Abnormal Breast Cancer Screening in an Urban Population: A Patient Navigation Intervention." *Cancer* 109: 359–67.

Brenner, H., A. Gondos, and V. Arndt. 2007. "Recent Major Progress in Long-Term Cancer Patient Survival Disclosed by Modeled Period Analysis." *Journal of Clinical Oncology* 25: 3274–80.

Campbell, M., R. Fitzpatrick, A. Haines, A. L. Kinmonth, P. Sandercock, D. Spiegelhalter, and P. Tyrer. 2000. "Framework for Design and Evaluation of Complex Interventions to Improve Health." *British Medical Journal* 321: 694–6.

Chang, S. W., K. Kerlikowske, A. Nápoles-Springer, S. F. Posner, E. A. Sickles, and E. J. Pérez-Stable. 1996. "Racial Differences in Timeliness of Follow-Up after Abnormal Screening Mammography." *Cancer* 78: 1395–402.

Dohan, D., and D. Schrag. 2005. "Using Navigators to Improve Care of Underserved Patients: Current Practices and Approaches." *Cancer* 104: 848–55.

Eccles, M., J. Grimshaw, M. Campbell, and C. Ramsay. 2003. "Research Designs for Studies Evaluating the Effectiveness of Change and Improvement Strategies." *Quality and Safety in Health Care* 12: 47–52.

Ell, K., B. Vourlekis, P. J. Lee, and B. Xie. 2007. "Patient Navigation and Case Management Following an Abnormal Mammogram: A Randomized Clinical Trial." *Preventive Medicine* 44: 26–33.

Ferrante, J. M., P.-H. Chen, and S. Kim. 2008. "The Effect of Patient Navigation on Time to Diagnosis, Anxiety, and Satisfaction in Urban Minority Women with Abnormal Mammograms: A Randomized Controlled Trial." *Journal of Urban Health* 85: 114–24.

Freeman, H. P. 2006. "Patient Navigation: A Community Centered Approach to Reducing Cancer Mortality." *Journal of Cancer Education* 21: S11–4.

Freund, K. M., T. A. Battaglia, E. Calhoun, D. J. Dudley, K. Fiscella, E. Paskett, P. C. Raich, R. C. Roetzheim, C. L. Bennett, J. A. Clark, R. Garcia, A. Greene, S. R. Patierno, and V. Warren-Mears. 2008. "National Cancer Institute Patient Navigation Research Program: Methods, Protocol and Measures." *Cancer* 113: 3391–9.

Gabram, S. G. A., M. J. B. Lund, J. Gardner, N. Hatchett, H. L. Bumpers, J. Okoli, M. Rizzo, B. J. Johnson, G. B. Kirkpatrick, and O. W. Brawley. 2008. "Effects of an Outreach and Internal Navigation Program on Breast Cancer Diagnosis in an Urban Cancer Center with a Large African–American Population." *Cancer* 113: 602–7.

Glaser, B. G., and A. L. Strauss. 1967. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago: Aldine Publishing Company.

Jandorf, L., A. Fatone, P. V. Borker, M. Levin, W. A. Esmond, B. Brenner, G. Butts, and W. H. Redd. 2006. "Creating Alliances to Improve Cancer Prevention and Detection among Urban Medically Underserved Minority Groups. The East Harlem Partnership for Cancer Awareness." *Cancer* 107: 2043–51.

Keating, N. L., J. Z. Ayanian, P. D. Cleary, and P. V. Marsden. 2007. "Factors Affecting Influential Discussions among Physicians: A Social Network Analysis of a Primary Care Practice." *Journal of General Internal Medicine* 22: 794–8.

Marsden, P. V. 1990. "Network Data and Measurement." *Annual Review of Sociology* 16: 435–63.

Norros, L., and M. Nuutinen. 2002. "The Concept of the Core Task and the Analysis of Working Practices." In *Work Process Knowledge*, edited by N. Boreham, R. Samurçay, and M. Fischer, pp. 25–39. London: Routledge.

Peterson, N. B., J. Han, and K. M. Freund. 2003. "Inadequate Follow-Up for Abnormal Pap Smears in an Urban Population." *Journal of the National Medical Association* 95: 825–32.

ResearchWare Inc. 2008. *HyperRESEARCH (Release 2.8) Qualitative Data Analysis Software*. Randolph, MA: ResearchWare Inc.

Ries, L. A. G., D. Melbert, M. Krapcho, A. Mariotto, B. A. Miller, E. J. Feuer, L. Clegg, M. J. Horner, N. Howlader, M. P. Eisner, M. Reichman, and B. K. Edwards editors. 2007. *SEER Cancer Statistics Review, 1975–2004.* Bethesda, MD: National Cancer Institute [accessed on March 24, 2008]. Available at http://seer.cancer.gov/csr/1975_2004/, based on November 2006 SEER data submission, posted to the SEER website, 2007.

Shavers, V. L., P. Fagan, and P. McDonald. 2007. "Health Disparities across the Cancer Continuum." *Journal of Health Care for the Poor and Underserved* 18: 1–5.

Smedley, B. D., A. Y. Stith, and A. R. Nelson, Committee on Understanding and Eliminating Racial and Ethnic Disparities in Health Care, Board on Health Sciences Policy, Institute of Medicine. 2003. *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care.* Washington, DC: National Academy of Sciences.

Turkle, S. 2003. "Sociable Technologies: Enhancing Human Performance When the Computer Is Not a Tool but a Companion." In *Converging Technologies for Improving Human Performance*, edited by M. C. Roco and W. S. Bainbridge, pp. 150–8. Dordrecht, the Netherlands: Kluwer Academic Publishers.

Vargas, R. B., G. W. Ryan, C. A. Jackson, R. Rodriguez, and H. P. Freeman. 2008. "Characteristics of the Original Patient Navigation Programs to Reduce Disparities in the Diagnosis and Treatment of Breast Cancer." *Cancer* 113: 426–33.

Weir, H. K., M. J. Thun, B. F. Hankey, L. A. Ries, H. L. Howe, P. A. Wingo, J. Ahmedin, E. Ward, R. N. Anderson, and B. K. Edwards. 2003. "Annual Report to the Nation on the Status of Cancer, 1975–2000, Featuring the Uses of Surveillance Data for Cancer Prevention and Control." *Journal of the National Cancer Institute* 95: 1276–99.

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.

Research Article

Statistics in Medicine

Received 17 April 2009, Accepted 20 April 2010 Published online 16 June 2010 in Wiley Interscience

(www.interscience.wiley.com) DOI: 10.1002/sim.3979

Approximate models for aggregate data when individual-level data sets are very large or unavailable

Erol A. Peköz,[a*†] Michael Shwartz,[a,b] Cindy L. Christiansen[c] and Dan Berlowitz[d]

In this article, we study a Bayesian hierarchical model for profiling health-care facilities using approximately sufficient statistics for aggregate facility-level data when the patient-level data sets are very large or unavailable. Starting with a desired patient-level model, we give several approximate models and the corresponding summary statistics necessary to implement the approximations. The key idea is to use sufficient statistics from an approximate model fitted by matching up derivatives of the models' log-likelihood functions. This derivative matching approach leads to an approximation that performs better than the commonly used approximation given in the literature. The performance of several approximation approaches is compared using data on 5 quality indicators from 32 Veterans Administration nursing homes. Copyright © 2010 John Wiley & Sons, Ltd.

Keywords: approximate Bayesian models; confidential data; Poisson binomial

1. Introduction

The field of approximate Bayesian computation (ABC) has developed in reaction to the computational demands of the growing size of data sets used in biology [1]. It is often necessary to summarize a huge amount of data by a few well-chosen summary statistics in situations where there are no exact sufficient statistics available. The recent article by Joyce and Marjoram [2] develops a framework for assessing whether an approximately sufficient statistics is useful; Le Cam [3] also addresses the concept of approximate sufficiency.

In this article, we look at a common statistical model often used to profile health-care facilities and develop customized approximately sufficient statistics. The setting of the problem is that there is a large amount of patient-level data located at many different facilities. A research group has developed risk-adjustment models to compute from the individual-level patient data a predicted probability that each patient experiences each of the different adverse events of interest and has developed a software that allows individual providers to calculate expected rates of adverse events; providers can compare these expected rates to observed rates as a way of monitoring their performance over time. In addition, a policy group would like to profile facilities. However, due to the volume of data and possible concerns about patient confidentiality, it is not feasible to transmit patient-level data to the policy group.

A realistic context for this problem arises out of software developed by the Agency for Healthcare Research and Quality (AHRQ) that hospitals can download in order to calculate risk-adjusted rates of patient safety indicators and of inpatient quality indicators from their own administrative data. As other measures of quality become acceptable (e.g. hospital re-admission rates), AHRQ is likely to continue to develop the software for calculation of relevant risk-adjusted rates. This software makes it easy for hospitals to use the quality indicators as part of their quality improvement programs. A useful service that AHRQ, other federal or state agencies, or private organizations could provide is to profile hospitals

[a]*Boston University School of Management, 595 Commonwealth Avenue, Boston, MA 02215, U.S.A.*
[b]*Center for Organization, Leadership and Management Research, VA Boston Healthcare System, Boston, MA, U.S.A.*
[c]*Boston University School of Public Health, 715 Albany Street, T-3W, Boston, MA 02118, U.S.A.*
[d]*Center for Health Quality, Outcomes and Economic Research, Bedford VA Hospital, 200 Springs Road, Bedford, MA 01730, U.S.A.*
[*]*Correspondence to: Erol A. Peköz, Boston University School of Management, 595 Commonwealth Avenue, Boston, MA 02215, U.S.A.*
[†]*E-mail: pekoz@bu.edu*

Contract/grant sponsor: Department of Veterans Affairs Health Services Research and Development Services; contract/grant number: IIR 06-260

based on the data generated from the software. Such an organization would encourage hospitals to submit results from use of the risk-adjustment software (much as AHRQ encourages hospitals to submit the results from use of its Hospital Survey on Patient Safety Culture, from which it then prepares comparative reports). To encourage flexibility in profiling, it would be most desirable for each hospital to submit, for each eligible patient for each adverse event of interest, the predicted probability of the outcome and whether or not the adverse event occurred. However, this would involve the transfer of large volumes of information. In addition, because patient-level information is being transferred, a variety of IRB considerations are likely to arise.

The goal of the research we present here is to identify useful summary statistics on quality indicators at each facility that can be conveniently transmitted to the profiling organization so that estimation approaches involving 'shrinkage' across facilities can be done. It is natural to consider sufficient statistics for this purpose, but there are none in our setting. We evaluate several alternate approaches to deal with the lack of sufficient statistics and identify one that performs well and does not introduce any computational complexity.

The organization of this article is as follows: In Section 2, we describe the statistical model and discuss the fact that sufficient statistics are not available. In Section 3, we describe the data to which we will apply our approach. In Section 4, we discuss two approaches for fitting approximate models for the data. In Section 5, we discuss the approximate models, in Section 6 we illustrate our approach using the data, in Section 7 we give a summary of the results, and Section 8 is an appendix containing the proofs.

2. The model

Let X_{ij} be a binary data variable that equals 1 if a particular adverse event happens to patient i, $1 \leqslant i \leqslant N_j$, in health-care facility j, $1 \leqslant j \leqslant M$, and it equals zero otherwise. We are interested in the following hierarchical model, which we subsequently refer to as the 'exact' model, where X_{ij} are the data values, p_{ij} are known constants, f is a given (known) link function and θ_j, μ, σ are unknown parameters that we would like to estimate:

$$X_{ij}|\theta_j, p_{ij} \sim \text{Bernoulli}(f(\theta_j, p_{ij})) \quad \text{and} \quad \theta_j|\mu, \sigma \sim \text{Normal}(\mu, \sigma^2). \tag{1}$$

In this model, p_{ij} is a risk-adjusted probability that has been previously calculated by taking into account various patient-specific characteristics, and it represents the chance patient i at facility j would have an adverse event if they were at an average facility. The parameter θ_j is a facility-specific factor, representing the quality at facility j, that increases or decreases the probability of an adverse event for the facility's patients. Note that this is a random effect model and therefore there should be shrinkage across facilities when estimating the model parameters. Our goal is to estimate the parameters θ_j for a function f such as

$$f(\theta, p) = \text{logit}^{-1}(\theta + \text{logit}(p)) = (1 + (1-p)e^{-\theta}/p)^{-1},$$

which we refer to as the logit link function, or

$$f(\theta, p) = \min(1, pe^{\theta}),$$

which we refer to as the log link function. These functions are both sensible to use, as they both equal p when $\theta = 0$ and they are increasing in θ. This means $\theta = 0$ corresponds approximately to an average facility and higher values of θ represent worse facilities. Typically, the function f will be increasing in θ and will have $f(0, p) = p$.

The standard Bayesian hierarchical modeling approach is to put prior distributions on the unspecified parameters and estimate the means of all the parameters conditional on the data. The difficulty in this situation is that the values of X_{ij} and p_{ij} are both confidential and are too numerous to conveniently transmit from each of the facilities to the main organization that performs the profiling. We need a method for summarizing these sets of numbers so that each facility only needs to report a small number of summary statistics. The usual approach is to use sufficient statistics, but, as we discuss in Section 4 below, there are no sufficient statistics we can use because the p_{ij} parameters are different across the N_j patients—if they were identical, we could use the total number of adverse events as a sufficient statistic which, conditional on the parameters, would have a binomial distribution. When they are not identical, the sum has what is called a Poisson-binomial distribution: the sum of independent Bernoulli trials with different success probabilities.

To implement the model in (1), it may seem as though each facility could simply transmit the maximum likelihood estimate of θ_j, but once these are gathered together there would be no obvious way to accurately allow for 'shrinkage' of the estimators across facilities. In fact, we test the approach of having each facility transmit the posterior mean and standard deviation of θ_j using separate (fixed-effect) models and then 'shrinking' them at the central organization using the assumption that they have a normal distribution. This approach does not work well on the data to which we applied it;

we call this the 'two-stage normal' approximation and illustrate it in Table III below. What is really necessary is for each facility to transmit enough information so that the central organization can, as best as possible, re-create the entire likelihood function for the data from each facility as a function of θ.

As there are no sufficient statistics available, the key idea we propose in this article is to use sufficient statistics from an approximate model fitted by matching moments or derivatives of the models' log-likelihood functions. In our recommended approach, described in Proposition 2 below, we use a binomial distribution to approximate the sum of non-identically distributed Bernoulli random variables where the two parameters of this distribution (the number of trials and the probability of success) are fitted using the first two moments of the probabilities from each facility. As a result, each health-care facility will simply need to report the total number of adverse events, as well as the first two moments of the predicted probabilities. For a different approach to approximating a similar model, see [4].

3. About the data

The data for this study were originally collected in 1998 from 35 Department of Veterans Affairs (VA) nursing homes that were selected to represent a balanced sample of different sizes, locations, and quality of care [5]. Complete data were available from 32 of the 35 nursing homes on five binary quality indicators (QIs) that reflect changes in patients' status over time. These 32 facilities are the ones included in our study. All the QIs have been used previously as measures of nursing home quality: pressure ulcer development [6–9]; functional decline [7–9]; behavioral decline [7, 10]; mortality [7, 11]; and preventable hospitalizations [12].

Data used in calculating QIs were from semi-annual patient assessments performed for case-mix-based reimbursements. Pressure ulcer development was recorded if a patient who was ulcer free at one assessment had a stage 2 or deeper pressure ulcer at the subsequent assessment. Functional decline was measured by a change between assessments in a score measuring limitations in eating, toileting, and transferring. Behavioral decline was measured by a change in a score measuring extent of verbal disruption, physical aggression, and socially inappropriate behavior. Mortality was recorded if there is a death within 6 months of an assessment regardless of location. Preventative hospitalizations occurred if the patient was admitted to an acute medical unit within 6 months of an assessment for one of 13 conditions identified as a potentially preventable hospitalization [13]. Risk-adjustment models have been developed for these QIs: pressure ulcer development [6], functional decline [14], behavioral decline and mortality [15], and preventable hospitalizations [16].

For each patient, the risk-adjustment models give a predicted probability of the adverse event in a six-month period based on the risk factors at the time of initial assessment. Thus, for each patient, we know whether each adverse event occurred (indicated by a 0 or 1) and the predicted probability of the adverse event from the relevant risk-adjustment model. Table I shows percentiles of the distribution of the predicted probabilities for each of the QIs. For 4 of the 5 QIs, individual predictions range from close to 0 to around 0.50 to 0.60; for mortality, there is a longer right tail, with predictions ranging almost to 1.

Table II shows percentiles of the distribution of the number of eligible cases (the total number of people who could potentially have the given adverse event) at each facility for the pressure ulcer and mortality QIs. The number of eligible cases ranges from 32 to over 400 for pressure ulcers (and for functional decline and behavioral decline, not shown in the table); and from slightly over 80 to almost 1200 for mortality (and preventable hospitalizations, also not shown). The second part of Table II shows percentiles of the distribution of the ratio of the observed number of adverse events at each facility to the predicted number for each of the QIs. The distribution is quite wide for 3 of the QIs, ranging from zero to over 2; it is tightest for mortality, ranging from 0.65 to 1.29.

Table I. Percentiles of the distribution of predicted probabilities by type of adverse event (pu—pressure ulcer development; fd—functional decline; bd—behavioral decline; mort—mortality; ph—preventable hospitalization).

Percentile	pu	fd	bd	mort	ph
Maximum	0.469	0.512	0.545	0.978	0.573
95th	0.107	0.348	0.320	0.759	0.197
90th	0.084	0.307	0.275	0.451	0.158
75th	0.055	0.250	0.206	0.221	0.109
50th	0.038	0.170	0.158	0.127	0.064
25th	0.021	0.081	0.127	0.074	0.027
10th	0.013	0.037	0.101	0.046	0.017
Minimum	0.011	0.011	0.062	0.006	0.004

Table II. Percentiles of the distribution of the number of eligible cases and of the ratio of observed to predicted number of adverse events in each of the 32 facilities (pu—pressure ulcer development; fd—functional decline; bd—behavioral decline; mort - mortality; ph—preventable hospitalization. The numbers of eligible cases for fd and bd are approximately the same as for pu; the number of eligible cases for ph is approximately the same as for mort).

	Number of eligible cases		Observed/predicted				
Percentile	pu	mort	pu	fd	bd	mort	ph
Minimum	32	83	0.00	0.15	0.00	0.65	0.23
25th	85	228	0.44	0.60	0.39	0.90	0.62
50th	159	324	0.90	0.96	0.81	0.99	0.96
75th	236	521	1.41	1.10	1.16	1.08	1.23
Maximum	408	1193	2.07	2.26	2.78	1.29	1.59

4. Two approaches for finding an approximate model

The log-likelihood function for the data at facility j under the model in (1) is

$$LL_j(\theta)=\sum_i X_{ij}\log f(\theta, p_{ij})+(1-X_{ij})\log(1-f(\theta, p_{ij})). \tag{2}$$

In general, there are no sufficient statistics for the individual-level data X_{ij} under this model. For example, with $M=1$ facility having $N_1=2$ patients with probabilities $p_1=\frac{1}{3}$, $p_2=\frac{1}{4}$ (omitting the subscript j) and $f(\theta, p)=\theta p$, we simplify (2) to get

$$LL(\theta)=X_1\log\frac{\theta}{3-\theta}+X_2\log\frac{\theta}{4-\theta}+\log\frac{\theta}{3-\theta}+\log\frac{\theta}{4-\theta}.$$

Since this log-likelihood is of the form aX_1+bX_2+a+b, it can be seen that this cannot be expressed in terms of two quantities where one depends on a, b but not X_1, X_2 and the other depends on X_1, X_2 but not a, b.

In the absence of a sufficient statistic, we therefore propose several alternative approximate models for the data for which the sum of the data values

$$X_j=\sum_i X_{ij}$$

is a sufficient statistic for a facility. Notice, we omit the subscript i to indicate that the binary variables have been summed over i for a given facility.

We say that an approximate model for the data is a k moment approximation if, when $\theta=0$, the first k moments of X_j conditional on all other parameters under the exact model coincide with the corresponding conditional moments under the approximate model. We also say that an approximate model for the data is a k derivative approximation for the exact model if the first k derivatives of its log-likelihood function with respect to θ evaluated at $\theta=0$ match the first k derivatives for the exact model $LL_j(\theta)$.

These definitions can be formalized by saying that an approximate model for the data X_j given θ_j along with some set of parameters S is a k moment approximation if, with the above definitions

$$E[(X_j)^m|\theta_j=0, p_{1j}, p_{2j}, \ldots]=E^a[(X_j)^m|\theta_j=0, S]$$

for $m=1, 2, \ldots, k$. We use the superscript a to denote expectations with respect to the approximate model. An approximate model with log-likelihood function $LL_j^a(\theta)$ for the data X_j is a k derivative approximation with respect to θ if

$$\frac{\partial^m}{\partial\theta^m}LL_j^a(\theta)|_{\theta=0}=\frac{\partial^m}{\partial\theta^m}LL_j(\theta)|_{\theta=0} \quad \text{for } m=1, 2, \ldots, k. \tag{3}$$

We say that an approximate model with multiple parameters is a k derivative approximation with respect to a given set of parameters if (3) holds with respect to each parameter separately.

Both these approaches seem sensible. The moment approximation is simpler in the sense that it does not depend on the link function f, whereas the derivative approximation does depend on f. As we show empirically later, when they are different, the derivative approximation performs better. For some choices of f the two approximations will coincide, whereas for others they will be different.

The intuition for why the derivative approximation performs better is the following: To estimate the distribution of X_j conditional on the parameters, it seems best to match moments to directly approximate the distribution. To approximate

the posterior distribution of θ conditional on the data requires the likelihood function and so it seems natural that the best performance is obtained when the likelihood function is directly approximated using an approach such as the derivative approach.

In order to find a good approximation for this model, it is important for practical reasons that it be easy to implement in standard Bayesian software packages like WINBUGS. Thus, it would be best to use only common distributions such as the binomial, Poisson or normal. In the article by Peköz *et al.* [17], several different approximations to the distribution of the sum of independent Bernoulli random variables are considered and evaluated. A Poisson approximation performs well if the Bernoulli probabilities are all very small. The normal approximation performs well if the probabilities are mostly near $\frac{1}{2}$. The binomial distribution performs well if the probabilities are either very small or very large overall—or very similar to each other. In practical applications, the probabilities will be quite diverse: some facilities will have very low numbers while others have probabilities spread more evenly between zero and one. It is therefore important to have an approximation that can flexibly adapt across all the situations. With a standard binomial approximation, it makes sense to pick the probability parameter to match the mean of the distribution to be approximated. One way of improving the binomial approximation is to let both the probability and the number of trials be two free parameters that can be fitted. This means in principle it may be possible to match the first two moments. Another idea for improving the Poisson approximation is to consider a translated Poisson approximation where a constant is added to a Poisson random variable. The constant and the Poisson rate are then two free parameters that can be adjusted to approximately match the first two moments.

A good combination of these ideas is to try a translated binomial approximation, where a constant is added to a binomial random variable. This then allows three parameters to be adjusted and, in most cases, three moments or three derivatives could be matched. The article by Peköz *et al.* [17] evaluates all the approximations discussed here (and derives error bounds) over a range of different settings and finds that a translated binomial approximation performs the best. As we show later, our analysis demonstrates that a binomial approximation with two fitted parameters performs much better than the usual binomial approximation with one fitted parameter, and the three parameter binomial approximation does not perform much better than the two parameter binomial approximation—though it requires additional computations and data to be transmitted and stored. We will therefore recommend the two-parameter binomial approximation, though we will illustrate all three approximations below. Once an approximation is fit to the data, the unknown parameters can be computed using either a Bayesian approach in WINBUGS, maximum likelihood estimation or other transform methods [18].

5. The approximate models

The first approximation we propose is one that is commonly used and is both a one-moment approximation and one-derivative approximation. It assumes that X_j, conditional on the parameters, has a binomial distribution, where the number of trials is the number of eligible cases and the binomial probability is chosen to be the average Bernoulli probability at that facility. To implement this approximation, each facility must simply transmit the total number of adverse events X_j along with the total number of eligible cases N_j and the average probability $p_j = \sum_i p_{ij}/N_j$ at that facility. This first approximation is summarized next.

Proposition 1
The one moment, one-derivative approximation. With the above definitions and the model in (1) with

$$f(\theta, p) = \text{logit}^{-1}(\theta + \text{logit}(p)),$$

and $p_j = \sum_i p_{ij}/N_j$, the model

$$X_j|\theta_j, p_j \sim \text{Binomial}(N_j, f(\theta_j, p_j))$$

is both a one-moment approximation and a one-derivative approximation with respect to θ_j, $j = 1, \ldots, M$ for the exact model (1).

Proof
See Appendix.

The second approximation we propose is both a two-moment approximation and a two-derivative approximation for the exact model. It assumes that X_j, conditional on the parameters has a binomial distribution, where the number of trials and the probability of success are both chosen to match the first two moments. This means that the number of trials in the binomial distribution n_j will not necessarily equal the number of eligible cases N_j in the facility. To implement this

approximation, each facility must simply transmit the total number of adverse events X_j along with the first two moments of the probabilities $\sum_i p_{ij}^2$ and $\sum_i p_{ij}$. We will recommend this approximation based on the numerical illustrations below.

Proposition 2
The two-moment, two-derivative approximation. With the above definitions and the model in (1) with

$$f(\theta, p) = \text{logit}^{-1}(\theta + \text{logit}(p)),$$

and $p_j = \sum_i p_{ij}^2 / \sum_i p_{ij}$, $n_j = \sum_i p_{ij} / p_j$, the model

$$X_j | \theta_j, n_j, p_j \sim \text{Binomial}(n_j, f(\theta_j, p_j))$$

is both a two-moment approximation and a two-derivative approximation with respect to θ_j, $j = 1, \ldots, M$ for the exact model (1).

Proof
See Appendix.

Remark 1
With the above definition, n_j may not necessarily be an integer, though the number of trials in a binomial distribution must be an integer. In principle, we could round off the number of trials to the nearest integer. In practice, we are using the likelihood as a function of θ and so the use of non-integer values in some sense represents an interpolation of the likelihood functions at the two adjacent integer numbers of trials. Standard software packages such as WINBUGS allow non-integer values for the number of trials in a binomial distribution and so a non-integer n_j is not a limitation of the approach. So that the propositions make sense, we can assume that n_j happens to be an integer—though the approximations can be improved using non-integer values.

Our third approximation is both a three-moment approximation and a three-derivative approximation. It assumes that X_j, conditional on the parameters, has a translated binomial distribution. A translated binomial distribution is simply a binomial distribution plus a constant. The three parameters, which consist of the number of trials, the probability of success and the amount translated, are chosen to match the first three moments. To implement this approximation, each facility must transmit the total number of adverse events X_j as well as the first three moments of the probabilities p_{ij}.

Proposition 3
The three-moment, three-derivative approximation. With the above definitions and the model in (1) with

$$f(\theta, p) = \text{logit}^{-1}(\theta + \text{logit}(p)),$$

and $\lambda_{kj} = \sum_i (p_{ij})^k$ along with

$$p_j = \frac{\lambda_{2j} - \lambda_{3j}}{\lambda_{1j} - \lambda_{2j}},$$

$$n_j = \frac{\lambda_{1j} - \lambda_{2j}}{p_j(1 - p_j)},$$

$$s_j = \lambda_{1j} - n_j p_j,$$

the model

$$(X_j - s_j | \theta_j, s_j, n_j, p_j) \sim \text{Binomial}(n_j, f(\theta_j, p_j))$$

is both a three-moment approximation and a three-derivative approximation with respect to θ_j, $j = 1, \ldots, M$ for the exact model (1).

Proof
See Appendix.

Remark 2
It may be that the quantity s_j is not an integer. However, as discussed in Remark 1 above, this is not a limitation.

Our next result is that for the log link function $f(\theta, p) = \min(1, pe^{\theta})$, the k moment approximations above do not coincide with the k derivative approximations. We can derive a different k derivative approximation. Empirical evidence shows that this performs much better than the k moment approximation.

Proposition 4
The two-derivative approximation differs from the two-moment approximation with the log link function. With the above definitions and the exact model in (1) with

$$f(\theta, p) = \min(1, pe^{\theta}),$$

and $p_j = \sum_i p_{ij}^2 / \sum_i p_{ij}$, $n_j = \sum_i p_{ij} / p_j$, the model

$$X_j | \theta_j, n_j, p_j \sim \text{Binomial}(n_j, f(\theta_j, p_j))$$

is a two-moment approximation but is not a two-derivative approximation with respect to θ_j, $j = 1, \ldots, M$ for the exact model (1).

Proof
See Appendix.

We next present the two-derivative model for the log link function.

Proposition 5
With the above definitions and the exact model in (1) with

$$f(\theta, p) = \min(1, pe^{\theta}),$$

and

$$\alpha_j = \sum_i \frac{p_{ij}}{1 - p_{ij}} (1 - X_{ij}),$$

$$\beta_j = \sum_i \frac{p_{ij}}{(1 - p_{ij})^2} (1 - X_{ij}),$$

and letting

$$p_j = 1 - \alpha_j / \beta_j,$$

and

$$n_j = \frac{\alpha_j^2}{\beta_j - \alpha_j} + X_j$$

the model

$$X_j | \theta_j, n_j, p_j \sim \text{Binomial}(n_j, f(\theta_j, p_j))$$

is a two-derivative approximation with respect to θ_j, $j = 1, \ldots, M$ for the exact model (1).

Proof
See Appendix.

The final result in this section is a straightforward extension to multiple types of adverse events. Note in this model that θ_j is an underlying latent measure of quality at facility j that is reflected in the data for each of the quality indicators.

Proposition 6
A model for multiple adverse events. Suppose there are m different types of adverse events. Let X_{ijk} equal 1 if person number i in facility number j has an adverse event of type k, $1 \leqslant k \leqslant m$, and let it equals zero otherwise. Let p_{ijk} be the given probability associated with this variable. For the kth type of adverse event, we define

$$f_k(\theta, p) = \text{logit}^{-1}(d_k + c_k \theta + \text{logit}(p)).$$

Suppose that

$$X_{ijk} | \theta_j, p_{ijk} \sim \text{Bernoulli}(f_k(\theta_j, p_{ijk})) \quad \text{and} \quad \theta_j | \mu, \sigma \sim \text{Normal}(0, 1) \tag{4}$$

is the exact model for some unknown parameters c_k, d_k that are to be estimated. With $X_{jk} = \sum_i X_{ijk}$ along with $p_{jk} = \sum_i (p_{ijk})^2 / \sum_i p_{ijk}$, $n_{jk} = \sum_i p_{ijk} / p_{jk}$, the model

$$(X_{jk} | \theta_j, n_{jk}, p_{jk}) \sim \text{Binomial}(n_{jk}, f_k(\theta_j, p_{jk}))$$

Table III. The approximate models.

Proposition	Approximation (link function)	Summary statistics necessary
1	1-moment, 1-derivative (logit)	$N_j, \sum_i p_{ij}, \sum_i X_{ij}, \forall j$
2	2-moment, 2-derivative (logit)	$\sum_i p_{ij}^2, \sum_i p_{ij}, \sum_i X_{ij}, \forall j$
3	3-moment, 3-derivative (logit)	$\sum_i p_{ij}^3, \sum_i p_{ij}^2, \sum_i p_{ij}, \sum_i X_{ij}, \forall j$
4	2-moment (log)	$\sum_i p_{ij}^2, \sum_i p_{ij}, \sum_i X_{ij}, \forall j$
5	2-derivative (log)	$\sum_i \frac{p_{ij}}{1-p_{ij}}(1-X_{ij}), \sum_i \frac{p_{ij}}{(1-p_{ij})^2}(1-X_{ij}), \sum_i X_{ij}, \forall j$
6	2-moment, multiple QIs (logit)	$\sum_i p_{ijk}^2, \sum_i p_{ijk}, \sum_i X_{ijk}, \forall j,k$

is both a two-moment approximation and a two-derivative approximation with respect to θ_j, $j=1,\ldots,M$ for the exact model in (4) when $c_k=1$, $d_k=0$ for $1\leqslant k\leqslant m$.

Proof
See Appendix.

We summarize all the approximations, along with the necessary summary statistics, in Table III.

6. Numerical results and comparison of the approaches

In this section, we illustrate the different approximations. We will see that the two-derivative approximations (for both link functions we consider) seem to perform much better than the one-derivative approximations, and (for the logit link function) the three-derivative approximation performs quite similarly to the two-derivative approximation—though all the approximations are fairly reasonable. We will illustrate the approximations using two different approaches. In the first approach, we test the approximations on data from the Veterans Health Administration—see Section 3 for a description of the data.

To do the computations, we put flat priors on the top level unknown parameters and then estimate the posterior means of the facility effects conditional on the data under the exact model and various approximate models. We use the notation $\bar{\theta}_j$ to denote our numerical estimate (computed by averaging together samples from the posterior distribution using the software package WINBUGS) of the posterior mean $E[\theta_j|X_{ij},\forall i,j]$ under the exact model and we use the notation $\bar{\theta}_j^a$ to denote our numerical estimate of the posterior mean $E^a[\theta_j|X_j,\forall j]$ under some given approximate model. For the illustrations below, we measure the accuracy of an approximation using the average of the absolute value of the difference between our estimate of the facility-effect parameters under the approximate and exact models. We compute this using

$$\text{Error}=\frac{1}{M}\sum_{j=1}^{M}|\bar{\theta}_j-\bar{\theta}_j^a| \tag{5}$$

except for the multiple quality indicators in Proposition 6, where we instead use

$$\text{Error}=\frac{1}{m}\sum_{k=1}^{m}\frac{1}{M}\sum_{j=1}^{M}|\bar{\theta}_{j,k}-\bar{\theta}_{j,k}^a|, \tag{6}$$

where $\theta_{j,k}=d_k+c_k\theta_j$ and $\theta_{j,k}^a=d_k^a+c_k^a\theta_j^a$.

First, we illustrate the approximations given in Propositions 1, 2, 3. The top half of Table IV gives the errors for models using the logit link function $f(\theta,p)=\text{logit}^{-1}(\theta+\text{logit}(p))$. The standard deviation of the effect sizes in Table V, when compared with the first row of Table IV, shows that the error from the one-moment approximation is not bad; in many cases, it is of a smaller order of magnitude than the estimated standard deviation (of the effect size θ in Table V). The second line shows that the two-moment approximation is always much better than the one-moment approximation. In several cases, the error is an order of magnitude improvement on the one-moment approximation in the prior line. The third line of the table shows that the three-moment approximation does not perform consistently better than the two-moment approximation with the logit link function. The rightmost column (labeled as '3QIs') shows that when multiple quality indicators are included in the model (using the approach of Proposition 6 above, where only the two-moment case is illustrated—the one- and three-moment versions are analogous to those in Propositions 1 and 3) the

Table IV. Errors associated with different approximations for the different quality indicators.

	pu	fd	bd	mort	ph	3QIs
Logit link function						
One-moment	0.015	0.028	0.021	0.042	0.020	0.016
Two-moment	0.004	0.003	0.007	0.003	0.002	0.004
Three-moment	0.004	0.003	0.007	0.002	0.004	0.004
Two-stage normal	0.103	0.018	0.182	0.003	0.012	
Log link function						
Two-moment	0.014	0.018	0.040	0.050	0.014	
Two-derivative	0.006	0.008	0.020	0.004	0.004	

Table V. Standard deviation of the effect size θ for the different quality indicators.

pu	fd	bd	mort	ph
0.420	0.465	0.982	0.123	0.422

two-moment approximation performs much better than the one-moment, and the three-moment approximation performs quite similarly to the two-moment approximation.

The fourth line of the table, labeled 'two-stage normal,' illustrates an approximation approach where each facility separately computes its best estimate of θ_j, which is then transmitted to the organization and 'shrunk' using a hierarchical model. Specifically, first the posterior mean and standard deviation for θ_j is computed separately for each facility using a fixed-effects model with a flat prior for θ_j. Then, these facility estimates are combined in a model where we assume the θ_j values are the sum of two draws from a normal distribution: one is a distribution with an unknown mean and standard deviation that is common to all facilities, and the other is an error term that is normally distributed with mean zero and standard deviation which is specific to each facility. The fourth line shows that this approximation approach performs better than the one-moment approximation in most cases—though in cases where there are facilities with zero observed adverse events (for pu and bd), it performs much worse. Some adjustments to the model would need to be made to account for zero values. But, the approximation is not nearly as good as the two-moment approach in most cases.

The last two rows of the table show, using the log link function $f(\theta, p)=\min(1, pe^{\theta})$, that the two-derivative approximation in Proposition 5 performs much better than the two-moment approximation. This is evidence for the theory that the derivative approach works better than the moment approach when they differ.

In the second approach, we use to illustrate the accuracy of the approximations, we generate simulated data and compare the maximum likelihood estimates of the unknown parameters across the different models. In particular, we create a single health-care facility with 100 patients having predetermined risk-adjusted probabilities. In the first simulation, we create these probabilities according to equally spaced percentiles of a beta distribution with a coefficient of variation equal to 1; in the second simulation, the coefficient of variation is equal to 0.4. We also vary the means of these distributions from 0 to 0.2. These ranges are chosen in order to mimic the summary statistics we saw in the actual data for the risk-adjusted probabilities. We also use different values of the unknown parameter θ ranging from 0 through 4. Given specific values of the parameters, we use the logit link function to generate simulated data and then compute the maximum likelihood estimate of θ under the exact model and then under the various approximations. Our simulations show that when θ is in the range from 0 to 2, the two-moment approximation generally performs better than the one-moment approximation. As the value of the θ parameter becomes very large (this corresponds to the case where there is a large amount of variation in the risk-adjusted probabilities across facilities and implies large differences in underlying quality), we see that a two-moment approximation performs worse than the one-moment approximation. As there is rarely such large variation in underlying quality across facilities, it is not a surprise that the two-moment approximation performs well when tested on real data.

7. Conclusion

In conclusion, we see that the two- and three-derivative approximations performed the best. The three-derivative approximation requires more data and performs similarly to the two-derivative approximation, and so we recommend the two-derivative approximation. For the logit link function, the two-derivative and two-moment approximations are identical though, for the log link function, the two-derivative approximation seems to perform better. To implement the

two-derivative approximation, each health-care facility needs to report three summary statistics to the central profiling organization: the first two moments of the predicted risk-adjusted probabilities, as well as the total number of adverse events occurring. Then, the two-parameter binomial distributional approximation described in Proposition 2 can be fitted and used to estimate the likelihood function.

The approach we recommend is nice because the approximation is quite simple and easy to implement. The standard (one-moment) binomial approximation is commonly used and this two-parameter binomial approximation is an improvement with essentially no additional computational costs or model complexity other than computing and transmitting one additional summary statistic.

Appendix

In this section, we prove the propositions above. We drop the facility subscript j in each proof.

Proof of Proposition 1
As the first moment of a binomial distribution equals np, the fact that this model is a one-moment estimator follows immediately. To see that it is also a one-derivative estimator, let

$$f(\theta, p)=\frac{1}{1+\dfrac{1-p}{p}e^{-\theta}},$$

and recall that the exact log-likelihood function is

$$LL(\theta)=\sum_i X_i \ln(f(\theta, p_i))+(1-X_i)\ln(1-f(\theta, p_i)).$$

The log-likelihood function for this approximation is

$$LL^a(\theta)=X\ln(f(\theta, p))+(N-X)\ln(1-f(\theta, p))+c,$$

where we write $X=\sum_i X_i$ and $c=\ln(N!/(X!(N-X)!)$. It then follows that

$$\frac{\partial}{\partial\theta}LL(\theta)=\sum_i(p_i e^{-\theta}-e^{-\theta}-p_i)^{-1}(p_i-X_i p_i-X_i e^{-\theta}+X_i p_i e^{-\theta})$$

along with

$$\frac{\partial}{\partial\theta}LL^a(\theta)=(pe^{-\theta}-e^{-\theta}-p)^{-1}(Np-Xp-Xe^{-\theta}+Xpe^{-\theta}),$$

and we obtain

$$\frac{\partial}{\partial\theta}LL(\theta)|_{\theta=0}=\sum_i(X_i-p_i),$$

and

$$\frac{\partial}{\partial\theta}LL^a(\theta)|_{\theta=0}=(X-Np).$$

Solving the equation

$$\frac{\partial}{\partial\theta}LL(\theta)|_{\theta=0}=\frac{\partial}{\partial\theta}LL^a(\theta)|_{\theta=0}$$

for p yields the result of the proposition. □

Proof of Proposition 2
As the mean and variance of a binomial distribution equals np and $np(1-p)$ respectively, the fact that this model is a two-moment estimator follows immediately. To see that it is also a two-derivative estimator, we again let

$$f(\theta, p)=\frac{1}{1+\dfrac{1-p}{p}e^{-\theta}},$$

and recall that the exact log-likelihood function is

$$LL(\theta)=\sum_i X_i \ln(f(\theta, p_i))+(1-X_i)\ln(1-f(\theta, p_i)).$$

The log-likelihood function for this approximation is

$$LL^a(\theta)=X\ln(f(\theta, p))+(n-X)\ln(1-f(\theta, p))+c,$$

where we write $X=\sum_i X_i$ and $c=\ln(N!/(X!(N-X)!)$ and now the parameters n, p are both to be solved for. It then follows, using the results from the previous proposition, that

$$\frac{\partial^2}{\partial\theta^2}LL(\theta)=\sum_i(p_i e^{-\theta}-e^{-\theta}-p_i)^{-2}(p_i-1)(e^{-\theta})p_i,$$

and

$$\frac{\partial^2}{\partial\theta^2}LL^a(\theta)=(pe^{-\theta}-e^{-\theta}-p)^{-2}(p-1)(e^{-\theta})np,$$

and we obtain

$$\frac{\partial^2}{\partial\theta^2}LL(\theta)|_{\theta=0}=\sum_i(p_i^2-p_i),$$

and

$$\frac{\partial^2}{\partial\theta^2}LL^a(\theta)|_{\theta=0}=(np^2-np).$$

Using the results from the previous proposition, we can then solve the system of equations

$$\frac{\partial}{\partial\theta}LL(\theta)|_{\theta=0}=\frac{\partial}{\partial\theta}LL^a(\theta)|_{\theta=0},$$

and

$$\frac{\partial^2}{\partial\theta^2}LL(\theta)|_{\theta=0}=\frac{\partial^2}{\partial\theta^2}LL^a(\theta)|_{\theta=0}$$

for both n and p to obtain the result of the proposition. □

Proof of Proposition 3

The mean, variance, and third-central moment of a binomial distribution equals np, $np(1-p)$ and $np(1-p)(1-2p)$, respectively, and for a Poisson-binomial distribution W with probabilities $p_1, p_2, \ldots$ and $\lambda_j=\sum_i(p_i)^j$, we have $E[W]=\lambda_1$, $\mathrm{Var}(W)=\lambda_1-\lambda_2$ and $E[(W-E[W])^3]=\lambda_1-3\lambda_2+2\lambda_3$. The fact that this model is a three-moment estimator follows from these facts. To see that it is also a three-derivative estimator, we again let

$$f(\theta, p)=\frac{1}{1+\dfrac{1-p}{p}e^{-\theta}},$$

and recall that the exact log-likelihood function is

$$LL(\theta)=\sum_i X_i \ln(f(\theta, p_i))+(1-X_i)\ln(1-f(\theta, p_i)).$$

The log-likelihood function for this approximation is

$$LL^a(\theta)=(X-s)\ln(f(\theta, p))+(n-X-s)\ln(1-f(\theta, p))+c,$$

where we write $X=\sum_i X_i$ and $c=\ln(N!/(X!(N-X)!)$ and now the parameters s, n, p are all to be solved for. It then follows, using the results from the previous proposition, that

$$\frac{\partial^3}{\partial\theta^3}LL(\theta)=\sum_i(p_i e^{-\theta}-e^{-\theta}-p_i)^{-3}(p_i-e^{-\theta}+p_i e^{-\theta})(p_i-1)(e^{-\theta})p_i,$$

and

$$\frac{\partial^3}{\partial\theta^3}LL^a(\theta)=(pe^{-\theta}-e^{-\theta}-p)^{-3}(p-e^{-\theta}+pe^{-\theta})(p-1)(e^{-\theta})np,$$

and we obtain

$$\frac{\partial^3}{\partial\theta^3}LL(\theta)|_{\theta=0}=\sum_i(3p_i^2-p_i-2p_i^3),$$

and

$$\frac{\partial^3}{\partial\theta^3}LL^a(\theta)|_{\theta=0}=(3np^2-np-2np^3).$$

Using the results from the previous proposition, we can then solve the system of equations

$$\frac{\partial}{\partial\theta}LL(\theta)|_{\theta=0}=\frac{\partial}{\partial\theta}LL^a(\theta)|_{\theta=0},$$

$$\frac{\partial^2}{\partial\theta^2}LL(\theta)|_{\theta=0}=\frac{\partial^2}{\partial\theta^2}LL^a(\theta)|_{\theta=0},$$

and

$$\frac{\partial^3}{\partial\theta^3}LL(\theta)|_{\theta=0}=\frac{\partial^3}{\partial\theta^3}LL^a(\theta)|_{\theta=0}$$

for s, n and p to obtain the result of the proposition. □

Proof of Propositions 4 and 5

The fact that this approximation is a two-moment approximation follows from the argument in the proof of Proposition 2 above. Let

$$f(\theta,p)=pe^{\theta}$$

and recall that the exact log-likelihood function is

$$LL(\theta)=\sum_i X_i\ln(f(\theta,p_i))+(1-X_i)\ln(1-f(\theta,p_i)).$$

The log-likelihood function for this approximation is

$$LL^a(\theta)=X\ln(f(\theta,p))+(n-X)\ln(1-f(\theta,p))+c,$$

where we write $X=\sum_i X_i$ and $c=\ln(N!/(X!(N-X)!)$ and now the parameters n, p are both to be solved for. It then follows that

$$\frac{\partial}{\partial\theta}LL(\theta)=\sum_i X_i-e^{\theta}p_i\frac{1-X_i}{1-e^{\theta}p_i},$$

$$\frac{\partial^2}{\partial\theta^2}LL(\theta)=\sum_i -e^{\theta}p_i\frac{1-X_i}{1-e^{\theta}p_i}-e^{2\theta}p_i^2\frac{1-X_i}{(1-e^{\theta}p_i)^2}$$

along with

$$\frac{\partial}{\partial\theta}LL^a(\theta)=X-pe^{\theta}\frac{n-X}{1-pe^{\theta}},$$

and

$$\frac{\partial^2}{\partial\theta^2}LL^a(\theta)=-pe^{\theta}\frac{n-X}{1-pe^{\theta}}-p^2e^{2\theta}\frac{n-X}{(1-pe^{\theta})^2},$$

and we obtain

$$\frac{\partial}{\partial\theta}LL(\theta)|_{\theta=0}=\sum_i -p_i\frac{1-X_i}{1-p_i}-p_i^2\frac{1-X_i}{(1-p_i)^2},$$

and

$$\frac{\partial}{\partial\theta}LL^a(\theta)|_{\theta=0}=X-p\frac{n-X}{1-p},$$

and

$$\frac{\partial^2}{\partial\theta^2}LL^a(\theta)|_{\theta=0}=-p\frac{n-X}{1-p}-p^2\frac{n-X}{(1-p)^2}.$$

We can then solve the system of equations

$$\frac{\partial}{\partial\theta}LL(\theta)|_{\theta=0}=\frac{\partial}{\partial\theta}LL^a(\theta)|_{\theta=0}$$

along with

$$\frac{\partial^2}{\partial\theta^2}LL(\theta)|_{\theta=0}=\frac{\partial^2}{\partial\theta^2}LL^a(\theta)|_{\theta=0}$$

for both n and p to obtain the result of the proposition. □

Proof of Proposition 6
This proposition follows immediately from the same argument used to prove Proposition 2. □

Acknowledgements

This research was supported in part by a grant from the Department of Veterans Affairs Health Services Research and Development Service, IIR 06-260. The authors thank the referees for many detailed suggestions that have improved the presentation of this article.

References

1. Plagnol V, Tavare S. Approximate Bayesian Computation and MCMC 2002. In *Monte Carlo and Quasi-Monte Carlo Methods,* Niederreiter H (ed.) Springer: New York, 2004; 99–114.
2. Joyce P, Marjoram P. Approximately sufficient statistics and Bayesian computation. *Statistical Applications in Genetics and Molecular Biology* 2008; **7**:Article 26.
3. Le Cam L. Sufficiency and approximate sufficiency. *Annals of Mathematical Statistics* 1964; **35**:1419–1455.
4. Miller S, Bradlow E, Dayaratna K. Closed-form bayesian inferences for the logit model via polynomial expansions. *Quantitative Marketing and Economics* 2006; **4**:173–206.
5. Berlowitz DR, Young GJ, Hickey EC, Saliba D, Mittman BS, Czarnowski E, Simon B, Anderson JJ, Ash AS, Rubenstein LV. Quality improvement implementation in the nursing home. *Health Services Research* 2003; **38**:65–83.
6. Berlowitz DR, Ash AS, Brandeis GH, Brand HK, Halpern JL, Moskowitz MA. Rating long-term care facilities on pressure ulcer development: importance of case-mix adjustment. *Annals of Internal Medicine* 1996; **124**:557–563.
7. Porell F, Caro FG, Silva A, Monane M. A longitudinal analysis of nursing home outcomes. *Health Services Research* 1998; **33**:835–865.
8. Mukamel DB. Risk-adjusted outcome measures and quality of care in nursing homes. *Medical Care* 1997; **35**:367–385.
9. Zimmerman DR, Karon SL, Arling G, Clark BR, Collins T, Ross R, Sainfort F. Development and testing of nursing home quality indicators. *Health Care Financing Review* 1995; **16**:107–127.
10. Arling G, Karon SL, Sainfort F, Zimmerman DR, Ross R. Risk adjustment of nursing home quality indicators. *Gerontologist* 1997; **37**:757–766.
11. Braun BI. The effect of nursing home quality on patient outcome. *Journal of the American Geriatric Society* 1991; **39**:329–338.
12. Carter MW. Factors associated with ambulatory care-sensitive hospitalizations among nursing home residents. *Journal of Aging and Health* 2003; **15**:295–330.
13. Agency for Healthcare Research and Quality. Prevention Quality Indicators Overview, 2004. Available from: http://www.qualityindicators.ahrq.gov/pqi_overview.htm. [accessed November 4, 2008].

14. Rosen A, Wu J, Chang B, Berlowitz DB, Rakovski C, Ash AS, Moskowitz MA. Risk adjustment for measuring health outcomes: an application in VA long-term care. *American Journal of Medical Quality* 2001; **16**:118–127.
15. Berlowitz DR, Rosen AK, Wang F, Tsilimingras D, Tariot PN, Kader B, Mukamel DB. Purchasing or providing nursing home care: can quality of care data provide guidance. *Journal of the American Geriatric Society* 2005; **53**:603–608.
16. Pizer SD, Wang M, Comstock C. Preventable hospitalization as a measure of quality of care in nursing homes. *Health Care Finance and Economics*, VA Bedford, MA. 2003; *Working Paper #2003-01*.
17. Peköz E, Röllin A, Čekanavičius V, Shwartz M. A three-parameter binomial approximation. *Journal of Applied Probability* 2009; **46**(4):1073–1085.
18. Brown LD, Cai T, Zhou H. Nonparametric regression in exponential families. *Technical Report*, 2008. Available from: http://www-stat.wharton.upenn.edu/~tcai/paper/html/VST-Exp-Regression.html [accessed April 16, 2009].

IJSEP, 2010, 8, 36-47

The Effect of a Cognitive Behavioral Exercise Intervention on Clinical Depression in a Multiethnic Sample of Women With Breast Cancer: A Randomized Controlled Trial

Frank M. Perna[1], Lynette Craft[2], Karen M. Freund[3], Gary Skrinar[4], Michael Stone[3], Lisa Kachnic[6], Carolyn Youren[3], Tracy A. Battaglia[3]

[1] Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, MD
[2] Department of Preventive Medicine, School of Medicine, Northwestern University, Chicago, IL
[3] Women's Health Unit, Section of General Internal Medicine, Boston University School of Medicine, Boston, MA
[4] Department of Health Sciences, Boston University, Boston, MA
[5] Department of Surgery, Boston University School of Medicine, Boston, MA
[6] Department of Radiation, Boston University School of Medicine, Boston, MA

Abstract

Exercise is known to facilitate physical and emotional adjustment among women treated for breast cancer, and exercise exerts a profound effect on clinical depression. However, the effect of exercise on reducing clinical depression among breast cancer patients has not been demonstrated, especially among ethnic minority women who have a higher incidence of depression and higher mortality following breast cancer. First, literature is presented to assess exercise effects on depression among women with breast cancer. Second, we present the results of a randomized controlled trial assessing the effect of a structured exercise intervention on depression and exercise behavior in a multiethnic sample of women with early stage breast cancer enrolled prior to the start of adjuvant treatment. Results suggest that, in comparison to population norms, the rate of depression was higher in breast cancer patients. Analyses further showed that the intervention significantly increased self-reported exercise and reduced depression. These data suggest that the beneficial effects of exercise may extend to breast cancer patients with depression and may be initiated prior to and during cancer treatment.

Key words: Exercise, breast cancer, depression

Corresponding author: Frank M. Perna, National Cancer Institute, Division of Cancer Control & Population Sciences, 6130 Executive Blvd., EPN 4074, Bethesda, MD 20892-7236. E-mail: pernafm@mail.nih.gov

Breast cancer is the second leading cause of cancer death for women in the United States, and the incidence rate of breast cancer has risen 24% in the last 10 years (Bigby & Holmes, 2005). It is estimated that clinical depression affects approximately 10-30% of women with early stage breast cancer, and patients diagnosed with cancer are more than three times as likely to experience depression in the following year (Deshields, Tibbs, Fan, & Taylor, 2006; Honda & Goodwin, 2004). Although this rate is higher than typically reported for women in the general population (20%), depression fortunately subsides over the course of breast cancer treatment and thereafter (Deshields et al., 2006). For example, Deshields and colleagues (2006) examined patterns of clinical depression among early stage breast cancer patients and noted that depression subsided in many women (recovery pattern) and did not occur in the majority of patients (resilience pattern), which is consistent with reports of positive emotional trajectories in most women in the year following surgery (Helgeson, Snyder, & Seltman, 2004).

Unfortunately, a significant number (12-29%) of women with breast cancer remain chronically depressed or become depressed during treatment and thereafter (Deshields et al., 2006). If left untreated, depression can become chronic and can increase the risk of chronic conditions such as cardiovascular disease (CVD), obesity, and diabetes (Chapman, Perry, & Strine, 2005). These data are particularly salient since many women—particularly black women—who are successfully treated for early stage breast cancer still experience untimely deaths attributable to CVD-related diseases (i.e., hypertension, diabetes; Bigby & Holmes, 2005).

Exercise and Clinical Depression

In the general population, exercise is consistently associated with a large reduction in clinical depression with effect sizes (ESs) ranging from 0.72 to 1.4 (Craft & Landers, 1998; Lawlor & Hopker, 2001). These ESs correspond to a 67-79% improvement in depression; only a 50% reduction in symptoms is considered clinically significant (Craft & Perna, 2004). Further, meta-analytic findings show the following regarding exercise: (a) it is effective for a variety of patient subgroups, (b) it is not significantly different from psychotherapy and pharmacological treatment, (c) it appears to equally benefit individuals with mild and severe depression (Craft & Landers, 1998).

Exercise and Breast Cancer

Randomized controlled trials have consistently demonstrated the psychosocial benefits of exercise among women with breast cancer; the added benefits of relieving or inhibiting cancer treatment-associated side effects (e.g., nausea and fatigue); and the improvement of cardiorespiratory fitness, body composition, and insulin sensitivity (Courneya, 2003; Galvao & Newton, 2005; Schmitz et al., 2005). Several of these physical health benefits have been proposed as mechanisms that may decrease the risk of recurrence and mortality following breast cancer treatment (Courneya et al., 2003; Demark-Wahnefried et al., 2001). Indeed, survival and morbidity following breast cancer was inversely related

to exercise frequency (Holmes, Chen, Feskanich, Kroenke, & Colditz, 2005). Moreover, the effect on mortality was demonstrated for relatively modest levels of exercise intensity and frequency, as demonstrated in walking programs. For these reasons, exercise has been advocated as a particularly important complimentary treatment for cancer patients (Courneya, 2003).

Breast Cancer and Depression Among Minority Women

The favorable effects of exercise notwithstanding, questions remain regarding the effectiveness of interventions for subsets of women with breast cancer. For example, although 5-year survival rates for breast cancer have improved significantly in the last decade, economically disadvantaged populations —Black women in particular—have relatively lower quality of life and 5-year survival rates for breast cancer, despite having a lower incidence rate (Bigby & Holmes, 2005; Bui, Ostir, Kuo, Freeman, & Goodwin, 2005). Research also suggests that despite similar depression rates between minority and Caucasian women, minority women are less likely to receive adequate treatment, to take antidepressants, or to receive specialty care (Dwight-Johnson, Sherbourne, Liao, & Wells, 2000; Miranda & Cooper, 2004). Further, minority women with breast cancer, who also have untreated depression, have more anxiety and pain, report poorer quality of life, and may die sooner (Ell et al., 2005; Goodwin, Zhang, & Ostir, 2004).

Exercise and Multiethnic Samples in Breast Cancer

Existing data suggest that interventions aimed at increasing the adoption of exercise are warranted, particularly for Black women with breast cancer who are at relatively greater mortality risk (Bigby & Holmes, 2005). However, a paucity of exercise trials exists. Those that have been conducted had participants who were mostly Caucasian (e.g., 80% or more) and resided in suburbia (Demark-Wahnefried et al., 2001; Mock et al., 1997; Mock et al., 2005; Pickett et al., 2002). Similarly, although small, uncontrolled, pilot exercise studies with Black women are promising (Wilson, Porter, Parker, & Kilpatrick, 2005) and the ES of exercise on depression across studies utilizing White samples is strong, there is a primary gap in the knowledge that centers on how well these findings generalize to depressed women of diverse racial and ethnic backgrounds. These data are especially important considering epidemiological findings that indicate the association of physical activity with reduced mortality is not moderated by race (Holmes et al., 2005; Schmitz et al., 2005).

The purpose of the present study was to assess the effect of a structured exercise intervention (SEI) in combination with cognitive behavioral exercise adherence counseling on depression and exercise behavior among a multiethnic sample of women with early stage breast cancer who were enrolled prior to the start of adjuvant treatment. This study extends the current literature by utilizing a prospective design, by examining depression prior to and during adjuvant treatment, and by recruiting a multiethnic and previously sedentary sample of women in an urban setting.

Method

Setting and Participants

The trial was conducted at the Boston University School of Medicine (BUSM) General Clinical Research Center and the Boston Medical Center (BMC). A joint BUSM and BMC institutional review board approved the study, and informed consent was obtained from all participants for all procedures. Eligibility criteria include (a) English speaking, (b) between 21 and 75 years of age, (c) sedentary lifestyle (i.e., exercise less than 3 times/week for greater than 30 min/session in last 6 months), (d) average or below average fitness as determined by a graded exercise test (GXT), and (e) recent diagnosis of breast cancer (Stage 0, I, II, or IIIa). Exclusion criteria include (a) non-cancer related contraindications to aerobic walking exercise (e.g., symptomatic coronary artery disease, psychotic spectrum mental illness, orthopedic problems), (b) pre-existing metabolic disease (e.g., diabetes, uncontrolled hypertension), and (c) a contraindication to exercise discovered on the exercise stress test (Winningham, 1991).

Experimental Design and Participant Recruitment

We present self-reported exercise and depression data on the 3-month follow-up of a larger study funded by the National Cancer Institute (CA R01-78801). The study employed a Group (structured intervention, information control) × Time (baseline, 3-month) prospective, randomized, and controlled clinical trial. Between April of 2001 and July of 2005, women who had completed breast cancer surgery but had not started radiation therapy, chemotherapy, or anti-hormone therapy were identified in weekly breast cancer case conferences. Post-doctoral fellows provided eligible breast cancer survivors with a study brochure and a consent form, and they discussed any questions. Baseline assessments consisted of a psychosocial survey, self-reported exercise, and physician-monitored GXT treadmill testing. Women in both intervention and control groups were given a $25 and $35 stipend for completing baseline and 3-month assessments, respectively.

Participants were stratified by cancer stage and randomly assigned to either a SEI or to an information control (IC) condition by use of a random number sequence table. Participant assignment to groups at enrollment was concealed from the project director, and the physicians monitoring GXTs were blinded to participant group assignment. Similarly, a physical therapist or an exercise physiologist, blinded to participant assignment, performed strength assessments.

Structured Exercise Intervention

Women assigned to the exercise group were given an individualized walking and resistance training program based on their GXT data, strength assessment, and established exercise recommendations for cancer patients (Winningham, 1991). Exercise training proceeded in two phases with a hospital-based portion followed by a transition to home-based exercise.

Hospital-based exercise. Participants were asked to come to exercise sessions three times a week for 4 weeks. For the aerobic component, participants began with 15-20 min of continuous treadmill walking 3 times per week at 50-70% of GXT-derived maximal heart rate (MHR). In subsequent weeks, duration was gradually increased by 5 min for a maximum of 40 min and a minimum of 30 min; intensity was increased to be within 70-85% of MHR according to participant comfort. Heart rate monitoring, rating of perceived exertion (RPE), and exercise enjoyment were used to assess participant comfort and to ensure participants were exercising in the appropriate training zone.

A resistance program consisted of upper body (i.e., bicep curl, triceps extension, chest fly, military press, upright row, and shoulder shrug) and lower body (i.e., leg squats and lunges) exercises. The upper-body resistance program was initiated with one set of 12 repetitions at the lightest weight, and as tolerated, repetitions were increased to 15 after the first week. After a participant could perform 15 repetitions of an exercise, another set was added. Upper-body exercises were performed with a padded weight belt with 1.0 lb bars that were used to adjust the total weight as necessary (range 0.5-20 lbs.). The participant's body weight was the only resistance used for lower body exercises. A physical therapist, kinesiologist, or exercise physiologist provided instruction on proper exercise technique. Participants were also provided with a manual, developed for this study, that used a woman model of the participants' approximate age and that depicted use of the adjustable weight belt, proper exercise technique, and key points for each exercise.

Home-based exercise. Participants were instructed to walk at least 3 times per week and were encouraged to walk most days of the week for 30 min or more. They were also provided with the adjustable weight belt and instructed to continue their resistance exercise 3 times a week. Participants were provided with monthly calendars to record their exercise activity and were contacted weekly by telephone or electronic mail according to patient preference.

Exercise adherence counseling and protocol delivery. Lastly, participants in the SEI condition received two 30-min exercise adherence counseling sessions during the hospital-based phase to enhance exercise maintenance for the transition to the home-based exercise. Participants were taught exertion cues and behavioral strategies (e.g. talk-test) to identify appropriate training zones and how to manage barriers encountered with a walking program (Winningham, 1991). The counseling also incorporated the use of behavioral contracting, contingency planning, goal setting, and self-monitoring techniques to promote exercise adherence (Prochaska & Marcus, 1994). Thereafter, SEI participants received weekly contact by telephone or electronic mail according to participant preference. To ensure uniformity of counseling, we developed a study-specific exercise adherence counseling manual for breast cancer and materials for staff training. The manual provided a detailed explanation of each counseling session that included the rationale, relevant behavior change, and micro-skills counseling principals; instructions to employ specific techniques (i.e., self-monitoring and goal setting); session handouts; and a checklist. For example, research assistant counselors specifically inquired regarding frequency and duration of exercise, instructed participants in the use of self-monitoring

calendars, provided verbal reinforcement and encouragement for exercise, and helped participants set goals. Lastly, as was appropriate, research staff aided participants in identifying solutions to existing or newly encountered exercise barriers and discussed ways to enlist social support for exercise. Clinical psychology and exercise psychology post-doctoral fellows delivered sessions and were supervised by a licensed psychologist with expertise in exercise and sport psychology.

Information Control

Women assigned to the IC condition received a 45-min information session consisting of individualized feedback of their physical assessment results and an informational brochure describing the potential physical and psychological benefits of exercise and their relevance to breast cancer. The brochure and information form describing the fitness component (e.g., oxygen consumption), the effect of exercise on the component, and the relevance for breast cancer were used to ensure uniform information delivery. However, the session specifically excluded the provision of an exercise prescription and discussion of strategies to address common exercise barriers. Participants who asked about an exercise prescription were told "do the best you can." To facilitate participant retention, control group participants were contacted once and then again 10 days prior to follow-up assessment.

Outcomes

The Center for Epidemiologic Studies Depression scale (CES-D) is a widely used and reliable (Cronbach Alpha $r = .85$) 20-item measure of depression that has been validated in breast cancer patients (Deshields et al., 2006). CES-D items indicative of depressive symptoms (e.g., "felt that I could not shake off the sad feelings even with help from my family or friends) are scored on a 0 (*Rarely or none of the time*) to 3 (*Most or all the time*) Likert scale. CES-D scores range from 0 to 60 and a cut-off score (≥ 16) to identify clinically significant depression has been established in both general and breast cancer samples.

Godin Leisure Time Exercise Questionnaire (LTEQ). The LTEQ self-report instrument surveys frequency of mild, moderate, and vigorous leisure-time physical activity over the past 7 days (Godin & Shephard, 1985). Metabolic equivalent values (3, 5, and 9 METS, respectively) are assigned to each exercise category and multiplied by frequency of occurrence to yield a total leisure score index. The LTEQ has been specifically validated in breast cancer studies, tested with ethnic minority populations, and has shown to be relatively free of social desirability (Courneya & Friedeneich, 1997; Courneya et al., 2003; Motl, McAuley, & DiStefano, 2005). The total leisure score is appropriately reliable (2-week test-retest reliability $r = .74$) and has been correlated with objective measures of fitness (accelerometer $r = .32$; VO_{2max} $r = .24$; Godin & Shepard, 1985; Motl et al., 2005). In our laboratory, we have validated LTEQ scores of breast cancer patients with VO_{2max} ($r = .28$, $p < .05$) and pedometer step counts ($r = .31$, $p < .05$).

Power Calculation and Data Analyses

Exercise intervention with breast cancer patients have large standardized effects (i.e., $d \geq .80$) on fitness and psychosocial outcomes (Courneya, 2003; Schmitz et al., 2005). An a priori power analysis, assuming a large effect ($d = .80$) and a .05 two-tailed alpha level, determined that 25 participants per group were needed to achieve .80 power. Fifty-seven participants were scheduled for baseline testing, and 7 participants were not enrolled due to undocumented hypertension, an orthopedic limitation, or abnormal blood pressure response to exercise testing. One participant was subsequently enrolled following a hypertension medication adjustment.

Analyses were conducted on all randomly assigned participants completing baseline assessments ($N = 51$). Because some participants missed follow-up assessments, and since intent-to-treat analyses are the recommended approach, we used several methods to impute missing values and then compared the concordance among results derived from each method (Tabachnick & Fidell, 2001). In all cases, these methods yielded the same result regarding interpretation of statistical significance. Because regression modeling is more widely applicable than carrying forward values, and since it is better than mean replacement, we used regression modeling to impute missing values to conduct our analyses.

Results

Participant Recruitment and Baseline Characteristics

As a first step, we used address and census data to compare eligible-enrolled to eligible-unenrolled participants and found they were not significantly different with respect to income, distance from BMC, and age ($p > .23$ for all calculations). We also had a high retention rate, with 80.4% randomly assigned participants completing follow-up assessments at 3 months; attrition was not significantly different between treatment and control groups at follow-up, $\chi^2(1, N = 51) = 1.45$, $p = .228$. Overall, participants completing follow-up assessments were not significantly different from follow-up noncompleters with respect to medical (i.e., cancer stage, surgery type, adjuvant treatment), demographic (i.e., age, race, marital status, employment, family income, or education), or exercise-related (i.e., BMI, VO_{2peak}, or arm and leg strength) characteristics.

We next compared intervention and control groups and found no significant differences with respect to demographic, cancer stage, treatment, and exercise characteristics. Participants ranged in age from 29-75 years ($M = 50.8$, $SD = 11.8$). Approximately one third (35%) worked full time and 37.3% were married or partnered. The majority (52.9%) of women had Stage I breast cancer and received lumpectomy surgery (74.1%). Many (44.1%) women received both radiation and chemotherapy, 26.5% received radiation only, 8.8% received chemotherapy only, and 20.6% received no adjuvant therapy. Women were generally obese (BMI: $M = 28.8$, $SD = 6.1$) and all had an average or below average fitness ($M = 21.8$, $SD = 6.3$ VO_{2peak} ml/kg/min). A large percentage of women were Black (44.1%), total ethnic minority group membership was high (45.1%), and about a third reported family incomes under $20,000.

Exercise Adherence and Protocol Delivery

Women assigned to the structured intervention completed an average of 83% of their scheduled hospital-based exercise sessions, (*M* = 9.9, *SD* = 3.3 sessions) and 76.9% completed all 12 sessions. These data are consistent with exercise adherence rates commonly reported in breast cancer literature (Courneya, 2003).

Depression

At baseline, 23.5% of women had CESD depression scores above the clinical cut-off, but depression rate did not differ between ethnic minority (26.1%) and non-minority women (21.4%) or between intervention (25.9%) and control (20.8%) groups, $\chi^2(1, N = 51) = .15$ and .18, $p = .699$ and .668, respectively. However, 33.3% of women were depressed at some point over the 3-month period, which was significantly higher than the general base rate of depression (e.g., approximately 20%) lifetime incidence in women, $\chi^2(1, N = 51) = 6.10$, $p = .014$.

Examination of intervention effect at 3-month follow-up revealed a significantly higher depression rate in the control group (37.5%) than in the intervention group (11.1%), $\chi^2(1, N = 51) = 4.917$, $p = .027$, $\varphi = -.311$. Analyzing the specific pattern of depression change indicated that 80% (4 of 5) of depressed participants remained depressed in the control group at 3-month follow-up, whereas only 42.9% (3 of 7) remained depressed in the intervention group. The rate of newly depressed patients (i.e., not depressed at baseline but depressed at follow-up) was 21% (4 of 19) in the control group, whereas it was 0% in the intervention group.

Exercise Behavior

At baseline, an ANOVA, with ethnic minority group and intervention group as independent factors and LTEQ score as a dependent factor, indicated that ethnic minority women were marginally less likely to exercise, $F(1, 47) = 3.63$, $p = .063$; however, control and intervention groups were not significantly different, $F(1, 47) = .19$, $p = .667$, and the interaction was nonsignificant, $F(1, 47) = .63$, $p = .43$. However at 3-month follow-up, a two-way ANCOVA, controlling for LTEQ score at baseline, revealed significantly higher LTEQ scores in the intervention group than in controls, $F(1, 47) = 9.99$, $p = .003$, $\eta^2 = .178$, but effects of ethnicity and the interaction between ethnicity and exercise-group membership were not significant, Fs (1, 47) = .11 and 1.67, ps = .739 and .203, respectively. These data suggest that the intervention was similarly effective in increasing self-reported exercise among ethnic minority and non-minority women. Specifically, LTEQ scores increased from baseline by 157% ($M = 9.7$, $SD = 8.1$ to $M = 25.0$, $SD = 13.1$) in the intervention group and by 32.7% among the control group ($M = 10.7$, $SD = 12.8$ to M=14.2, $SD = 11.8$).

Correlation Between Exercise and Depression

Finally, we sought to determine the relationship between exercise and depression level for those participants meeting criteria for clinical depression at any point during the study. We correlated LTEQ change with CESD change (from baseline to 3-month follow-up) and found a moderate but nonsignificant correlation of $r(17) = .38$, $p = .13$, (equivalent ES, $d = .69$).

Discussion

This is the first prospective, randomized controlled trial examining the exercise-depression relationship using intent-to-treat analyses in a multiethnic sample of breast cancer survivors. We found a similar depression rate between ethnic minority and non-minority women with breast cancer, but the overall depression rate (33.3%) was significantly higher than the lifetime population rate for women. These data are consistent with previous research with cancer patients (Deshields et al., 2006; Ell et al., 2005). Our results further demonstrated that SEI significantly improved physical activity and depression rate, and the intervention was similarly effective for ethnic minority and non-minority women. The amount of exercise had only a moderate, non-significant, correlation with exercise depression level.

These findings suggest that a structured intervention that transitions women from hospital- to home-based exercise may be an effective complimentary treatment to improve mental and physical health in breast cancer patients with depression. Moreover, the intervention ES (i.e., $d = -.69$) on depression we observed was comparable to the overall ES reported for depression among nonmedical population (i.e., $d = -.72$; Craft & Perna, 2004). These results are particularly important for minority women who are less likely to receive care for depression (Cooper et al., 2003), more likely than their Caucasian peers to be physically inactive (Crespo, Smit, Andersen, Carter-Pokras, & Ainsworth, 2000), have poorer breast cancer outcomes, and are at increased risk for CVD (Bigby & Holmes, 2005).

Assessment of the relationship between changes in physical activity and changes in depression revealed a moderate, but statistically nonsignificant, correlation. This may be attributed to several factors and study limitations.

One such limitation is the depression variable. Our study was not designed to specifically recruit depressed breast cancer patients, and thus, the subgroup correlation analysis' lack of statistical power may have contributed to non-significant findings. Similarly, although the CESD has established cut scores for depression, we did not verify depression by other means. However, a higher-than-expected rate of depression was observed in the sample that could not be attributed to other factors, such as support group attendance or antidepressant medication use.

Second, although not a study limitation per se, the lack of a significant inverse correlation between amount of exercise and depression may be questioned. It is possible that even a modest level of exercise is sufficient to promote positive changes in mental health among breast cancer patients and, once a threshold amount of exercise is met, no

further mental health benefit accrues from additional activity. This position is consistent with meta-analytic findings and with the magnitude of the correlation between exercise and depression we observed (Craft & Landers, 1998). These views are also consistent with prospective data from breast cancer patients indicating that improved mortality was achieved with only a minimal amount of exercise, whereas higher exercise level was not associated with improved mortality (Holmes et al., 2005). Given that many women with breast cancer may exercise with a lower frequency, and that this may still produce some physiological and psychological benefits, future research should manipulate exercise frequency and compare the effect of low-frequency exercise (e.g., 1-2 times per week) with more frequent exercise.

Third, we used a multifaceted intervention including a number of cognitive behavioral therapy characteristics. As such, we could not determine exactly which intervention component may have had the greatest effect. It is possible that intervention aspects, other than exercise, led to reductions in depression. For example, women in the SEI group engaged in self-monitoring and goal setting activities and may have derived a sense of self-efficacy and enhanced self-worth from meeting exercise goals. Similarly, women in the SEI group were contacted frequently to obtain data and were provided with brief supportive contact for their exercise attempts, and they may have perceived this support as therapeutic. Therefore, the exact mechanisms underpinning the intervention effect on depression are unclear; it is likely that various program components acted synergistically to increase exercise behavior and lessen depression among these women.

Finally, criteria for empirically validated therapies distinguish efficacious from specifically efficacious treatments (Chambless & Hollon, 1998). Treatment efficacy requires a finding that an intervention is superior to a control condition, whereas evidence that an intervention is superior to another treatment or an attention control is a criterion for treatment specificity. In the present study, research staff exposure time was not equivalent between conditions, but the design ensured that control and experimental groups received similar informational content, which was viewed as helpful, and the control condition represented an improvement over the standard no-treatment, control condition. As such, we have demonstrated the efficacy of an intervention to improve time spent in physical activity and to reduce the prevalence of depression among women with early stage breast cancer. Future research is needed to elucidate the mechanisms underlying this relationship.

References

Bigby, J., & Holmes, M. D. (2005). Disparities across the breast cancer continuum. *Cancer Causes Control, 16*(1), 35-44.

Bui, Q. U., Ostir, G. V., Kuo, Y. F., Freeman, J., & Goodwin, J. S. (2005). Relationship of depression to patient satisfaction: Findings from the barriers to breast cancer study. *Breast Cancer Research and Treatment, 89*(1), 23-28.

Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology, 66,* 7-18.

Chapman, D. P., Perry, G. S., & Strine, T. W. (2005). The vital link between chronic disease and depressive disorders [Electronic version]. *Preventing Chronic Disease, 2*(1), 1-10.

Cooper, L. A., Gonzales, J. J., Gallo, J. J., Rost, K. M., Meredith, L. S., Rubenstein, L. V., et al. (2003). The acceptability of treatment for depression among African-American, Hispanic, and White primary care patients. *Medical Care, 41,* 479-489.

Courneya, K. S. (2003). Exercise in cancer survivors: An overview of research. *Medicine & Science in Sports & Exercise, 35,* 1846-1852.

Courneya, K. S., & Friedeneich, C. M. (1997). Relationship between exercise during treatment and current quality of life among survivors of breast cancer. *Journal of Psychosocial Oncology, 15*(3), 35-57.

Courneya, K. S., Mackey, J. R., Bell, G. J., Jones, L. W., Field, C. J., & Fairey, A. S. (2003). Randomized controlled trial of exercise training in postmenopausal breast cancer survivors: Cardiopulmonary and quality of life outcomes. *Journal of Clinical Oncology, 21,* 1660-1668.

Craft, L. L., & Landers, D. M. (1998). The effect of exercise on clinical depression and depression resulting from mental illness: A meta-analysis. *Journal of Sport & Exercise Psychology, 20,* 339-357.

Craft, L. L., & Perna, F. M. (2004). The benefits of exercise for the clinically depressed. *Primary Care Companion to the Journal of Clinical Psychiatry, 6*(3), 104-111.

Crespo, C. J., Smit, E., Andersen, R. E., Carter-Pokras, O., & Ainsworth, B. E. (2000). Race/ethnicity, social class and their relation to physical inactivity during leisure time: Results from the Third National Health and Nutrition Examination Survey, 1988-1994. *American Journal of Preventive Medicine, 18,* 46-53.

Demark-Wahnefried, W., Peterson, B. L., Winer, E. P., Marks, L., Aziz, N., Marcom, P. K., et al. (2001). Changes in weight, body composition, and factors influencing energy balance among premenopausal breast cancer patients receiving adjuvant chemotherapy. *Journal of Clinical Oncology, 19,* 2381-2389.

Deshields, T., Tibbs, T., Fan, M. Y., & Taylor, M. (2006). Differences in patterns of depression after treatment for breast cancer. *Psychooncology, 15,* 398-406.

Dwight-Johnson, M., Sherbourne, C. D., Liao, D., & Wells, K. B. (2000). Treatment preferences among depressed primary care patients. *Journal of General Internal Medicine, 15,* 527-534.

Ell, K., Sanchez, K., Vourlekis, P., Lee, P. J., Dwight-Johnson, M., Langomasino, I., et al. (2005). Depression, correlates of depression, and receipt of depression care among low-income women with breast or gynecologic cancer. *Journal of Clinical Oncology, 23,* 3052-3060.

Galvao, D. A., & Newton, R. U. (2005). Review of exercise intervention studies in cancer patients. *Journal of Clinical Oncology, 23,* 899-909.

Godin, G., & Shephard, R. J. (1985). A simple method to assess exercise behavior in the community. *Canadian Journal of Applied Sport Sciences, 10*(3), 141-146.

Goodwin, J. S., Zhang, D. D., & Ostir, G. V. (2004). Effect of depression on diagnosis, treatment, and survival of older women with breast cancer. *Journal of American Geriatrics Society, 52*(1), 106-111.

Helgeson, V. S., Snyder, P., & Seltman, H. (2004). Psychological and physical adjustment to breast cancer over 4 years: Identifying distinct trajectories of change. *Health Psychology, 23*(1), 3-15.

Holmes, M. D., Chen, W. Y., Feskanich, D., Kroenke, C. H., & Colditz, G. A. (2005). Physical activity and survival after breast cancer diagnosis. *Journal of the American Medical Association, 293,* 2479-2486.

Honda, K., & Goodwin, R. D. (2004). Cancer and mental disorders in a national community sample: Findings from the national comorbidity survey. *Psychotherapy and Psychosomatics, 73,* 235-242.

Lawlor, D. A., & Hopker, S. W. (2001). The effectiveness of exercise as an intervention in the management of depression: Systematic review and meta-regression analysis of randomised controlled trials. *British Medical Journal, 322, 763-767.*

Miranda, J., & Cooper, L. A. (2004). Disparities in care for depression among primary care patients. *Journal of General Internal Medicine, 19,* 1-10.

Mock, V., Dow, K. H., Meares, C. J., Grimm, P. M., Dienemann, J. A., Haisfield-Wolfe, M. E., et al. (1997). Effects of exercise on fatigue, physical functioning, and emotional distress during radiation therapy for breast cancer. *Oncology Nursing Forum, 24*, 991-1000.

Mock, V., Frangakis, C., Davidson, N. E., Ropka, M. E., Pickett, M., Poniatowski, B., et al. (2005). Exercise manages fatigue during breast cancer treatment: A randomized controlled trial. *Psychooncology, 14*, 464-477.

Motl, R. W., McAuley, E., & DiStefano, C. (2005). Is social desirability associated with self-reported physical activity? *Preventive Medicine, 40*, 735-739.

Pickett, M., Mock, V., Ropka, M. E., Cameron, L., Coleman, M., & Podewils, L. (2002). Adherence to moderate-intensity exercise during breast cancer therapy. *Cancer Practice, 10*, 284-292.

Prochaska, J. O., & Marcus, B. H. (1994). The transtheoretical model: Applications to exercise. In R. K. Dishman (Ed.), *Advances in exercise adherence* (pp. 161-180). Champaign, IL: Human Kinetics.

Schmitz, K. H., Holtzman, J., Courneya, K. S., Masse, L. C., Duval, S., & Kane, R. (2005). Controlled physical activity trials in cancer survivors: A systematic review and meta-analysis. *Cancer Epidemiology, Biomarkers & Prevention, 14*, 1588-1595.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Boston: Allyn and Bacon.

Wilson, D. B., Porter, J. S., Parker, G., & Kilpatrick, J. (2005). Anthropometric changes using a walking intervention in African American breast cancer survivors: A pilot study. *Preventing Chronic Disease, 2*, A16.

Winningham, M. L. (1991). Walking program for people with cancer: Getting started. *Cancer Nursing, 12*, 219-225.

Acknowledgements

We wish to acknowledge the valuable contributions of our physicians to the project including Rita Blanchard, Michele, David, Elizabeth Dupuis, Maureen Kavanah, Renee McKinney, Jane Mendez, and Carol Rosenberg by Noreen O'Shea and the nursing staff of the General Clinical Research Center, by Lara Johnson, and most importantly, by our participant breast cancer survivors.

Author Note

This work was funded by the National Cancer Institute (CAR0178801), and by the National Institutes of Health, General Clinical Research Grant, (M01RR00533). This work was completed while Dr. Perna was at Boston University School of Medicine in the Women's Health Unit, Section of General Internal Medicine.

were boys who had disease durations range from 4 months to 2 years. Associated atopic conditions were found in four out of five patients.

Vision was impaired in all patients. Indeed, visual testing was difficult due to symptomatic itching and photophobia. Those with corneal complications did worse (20/100 in patients 1, 3 and 5, who had corneal disease). Patient 1 had an established shield ulcer which failed to respond to hourly topical steroid, autologous serum and regular debridement. Patient 5 had advanced glaucomatous field losses with cup–disc ratios of 0.9 bilaterally, bilateral pseudophakia and glaucoma-drainage-devices. She had an old corneal scar and a baseline vision of 20/100. To control symptomatic itching and superior giant papillae, she had been using topical cyclosporine 0.05% twice daily with only moderate relief.

Topical tacrolimus 0.03% ointment twice daily was introduced at various time points. It was used for the shield ulcer that failed to respond (patient 1) after 3 weeks of treatment. For patient 2, it was used in place for steroid and mast-cell-targeted treatment, which added up to 16 administrations per day. For patients 3 and 4, it was added in substitution for topical steroid therapy when corneal epithelial breakdown was first noted. In patient 5, it was used instead of topical cyclosporine, owing to its lack of efficacy.

All patients responded. The only side effect was burning sensation that did not require cessation of treatment. There was significant reduction in symptoms and signs. Giant papillae regressed as early as 1 week post-treatment. The large shield ulcer seen in patient 1 healed within 3 weeks. For the two cases with early epithelial disease, the defects healed within 1 week with scarring (figure 1). Total treatment duration ranged from 1 to 5 months. Symptoms redeveloped in two patients after 2 months, which was controlled with mast-cell-targeted treatments alone.

In summary, topical tacrolimus 0.03% was found to be safe and effective for VKC. We would suggest its early use as a steroid-sparing agent to prevent blinding complications.

P M K Tam, A L Young, L L Cheng, P T H Lam

Department of Ophthalmology and Visual Sciences, The Chinese University of Hong Kong, Prince of Wales Hospital, Shatin, Hong Kong, China

Correspondence to Alvin L. Young, Cornea and External Eye Disease, Department of Ophthalmology and Visual Sciences, The Chinese University of Hong Kong, Prince of Wales Hospital, Shatin, Hong Kong, China; youngla@ha.org.hk

Competing interests None.

Provenance and peer review Not commissioned; not externally peer reviewed.

Accepted 15 December 2009
Published Online First 16 June 2010

Br J Ophthalmol 2010;**94**:1405–1406.
doi:10.1136/bjo.2009.172387

REFERENCES

1. **Siddique M,** Manzouri B, Flynn TH, *et al*. Allergy and contact lenses. *Chem Immunol Allergy* 2007;**92**:166–75.
2. **Cetinkaya A,** Akova YA, Dursun D, *et al*. Topical cyclosporine in the management of shield ulcers. *Cornea* 2004;**23**:194–200.
3. **Schreiber SL,** Crabtree GR. The mechanism of action of cyclosporine A and FK506. *Immunol Today* 1992;**13**:136–42.
4. **Kymionis GD,** Goldman D, Ide T, *et al*. Tacrolimus ointment 0.03% in the eye for treatment of giant papillary conjunctivitis. *Cornea* 2008;**27**:228–9.
5. **Tam PM,** Young AL, Cheng LL, *et al*. Topical 0.03% tacrolimus ointment in the management of ocular surface inflammation in chronic graft-versus-host-disease. *Bone Marrow Transplant* 2009; advance online publication 5 Oct 2009.
6. **Attas-Fox L,** Barkana Y, Iskhakov V, *et al*. Topical tacrolimus 0.03% ointment for intractable allergic conjunctivitis: an open-label pilot study. *Curr Eye Res* 2008;**33**:545–9.

Ocular manifestations of torture: solar retinopathy as a result of forced solar gazing

Solar retinopathy as a result of sun gazing has been well documented and occurs as a result of thermal and photochemical processes after solar exposure.[1] Although solar exposure is usually a result of deliberate sun gazing, forced sun gazing can also be used as a torture method.

Torture is prevalent worldwide. In 2007, the Amnesty International documented cases of "torture and other cruel, inhuman or degrading treatment" in 81 countries.[2]

It is important for clinicians to be aware of the physical manifestations of torture because they are not always obvious. Torture methods are often devised to leave minimal long-term physical indications but often have tremendous psychological impact.[3] Survivors will often not disclose the experience of their torture with their physician, and the diagnosis may be missed.[4]

CASE

A 58-year-old West African man was referred to an ophthalmology clinic by his primary care physician with a chief complaint of difficulty reading and discomfort and blur on more than momentary reading. His medical history included torture in Cameroon 3 years earlier. On two consecutive days, he was taken from his cell and at gunpoint forced to stare at the sun for a duration that he believes was about an hour each day. He possessed no means for measuring time. He was threatened with death if he looked away or closed his eyes, and did observe the execution of another prisoner for failure to comply. He described great pain and blurred vision afterwards, but he felt that he recovered and was able to see well at close range and distance until recently when he developed difficulty reading.

On examination, the distance visual acuity was 20/20 in both eyes on the Snellen test without correction. Uncorrected near acuity was J7 in either eye. With appropriate presbyopic correction, his near vision improved to J0 in both eyes. Fundus examination of the left eye showed an atrophic area with pigment migration, shaped like two overlapping circles, approximately 1 mm in diameter. On retinal photographs, the lesion was exactly the same distance temporal to the fovea as the disc was nasal to the foveal umbo (see figure 1). The right fundus was normal. The visual fields were normal.

Diagnostic impressions were presbyopia *in both eyes* and solar burn in left eye secondary to torture by forced sun gazing. The appearance of the lesion is classic for the late lesion of solar gazing after the initial oedema of the neurosensory retina has faded. The patient was certain that he had looked directly at the sun, but the location of the unilateral burn suggested that he had instead instinctively arranged his gaze to place the image on the optic nerve head oculus dexter and temporal to the fovea *oculus sinister* to minimise pain. Because there was no optic atrophy, the lighter colour of the optic nerve presumably converted less light energy to heat than the darker fundus, including the retinal pigment epithelium and the choroidal pigment, and optic nerve tissue was not lost. In addition, the size of the extrafoveal lesion is comparable with the foveal lesions of patients seen with solar injury to the fovea.[5]

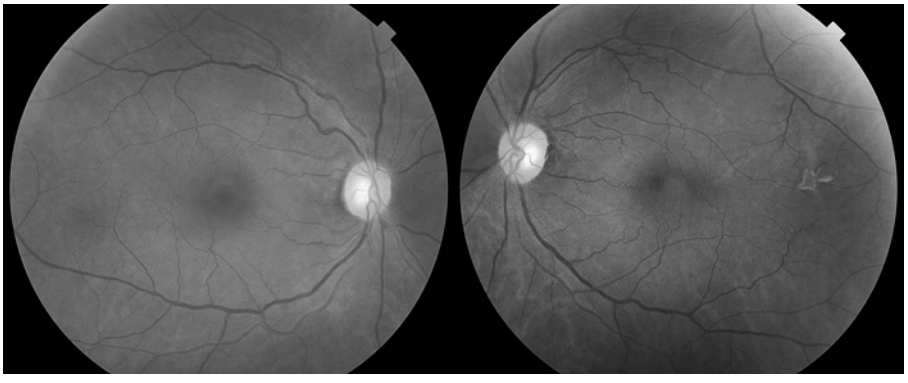


Figure 1 Fundi of patient showing scar temporal to the fovea in left eye secondary to solar burn.

Treatment with presbyopic correction eliminated his presenting symptoms.

DISCUSSION

In this report, we document that forced sun gazing can also be used as a physical torture method, which can result in solar retinopathy. Psychological sequelae may also be present in torture survivors, such as symptoms of post-traumatic stress disorder and major depression. If a physician identifies a patient as being a survivor of torture, it is important for them to be aware of programmes that specialise in the care of torture survivors such as the International Rehabilitation Council for Torture Victims.[6]

In summary, this is a unique report of forced sun gazing used as a torture method and resulting in solar retinopathy. Unlike with deliberate sun gazing, where solar burns are usually located foveal or perifoveal, torture victims may shift their eyes, causing solar burns offset from the retina. In the case presented in this report, an affidavit and testimony were given in court documenting the ophthalmic findings, which contributed to the patient being granted political asylum.

J Polat,[1] E Feinberg,[2] S S Crosby[3,4]

[1]University School of Medicine, Boston, Massachusetts, USA; [2]Department of Ophthalmology, Boston University School of Medicine, Boston, Massachusetts, USA; [3]Department of Medicine, Boston University School of Medicine, Boston, Massachusetts, USA; [4]Department of Health Law, Bioethics and Human Rights, Boston University School of Public Health, Boston, Massachusetts, USA

Correspondence to S S Crosby, Section of General Internal Medicine, Boston Medical Center, Crosstown 2, 801 Massachusetts Avenue, Boston, MA 02118, USA; scrosby@bu.edu

Competing interests None.

Patient consent Obtained.

Ethics approval This study was conducted with the approval granted by the institutional review board of the Boston University School of Medicine.

Provenance and peer review Not commissioned; externally peer reviewed.

Accepted 20 December 2009
Published Online First 24 June 2010

Br J Ophthalmol 2010;**94**:1406–1407.
doi:10.1136/bjo.2009.171595

REFERENCES

1. **Istock TH.** Solar retinopathy: a review of literature and case report. *J Am Optom Assoc* 1985;**56**:374–81.
2. **Amnesty International.** Amnesty International report 2008: at a glance. 2008:http://archive.amnesty.org/air2008/eng/facts-and-figures.html (accessed 30 Jun 2009).
3. **Iacopino V,** Allden K, Keller A. *Examining asylum seekers: a health and professional's guide to medical and psychological evaluations of torture*. Boston: Physicians for Human Rights, 2001.
4. **Miles SH,** Freedman AM. Medical ethics and torture: revising the declaration of Tokyo. *Lancet* 2009;**373**:344–8.
5. **Yannuzzi LA,** Fisher YL, Krueger A, *et al*. Solar retinopathy: a photobiological and geophysical analysis. *Trans Am Ophthalmol Soc* 1987;**85**:120–58.
6. International Rehabilitation Council for Torture Victims. http://www.irct.org/ (accessed 2 Jul 2009).

Decreased susceptibility to quinolones in methicillin-resistant *Staphylococcus aureus* isolated from ocular infections at a tertiary eye care centre

Methicillin-resistant *Staphylococcus aureus* (MRSA) is a serious cause of morbidity and mortality worldwide because of its multiple-drug resistance.[1] In the past, MRSA infections were considered as hospital acquired; however, in the 1990s, serious MRSA infections were reported in patients with no previous contact with the healthcare system.[2] Aggressive infections due to MRSA were observed in the eye and orbit in patients with no hospital exposure.[3] There are very few reports on ocular MRSA infections, and to the best of our knowledge, there are no reports on MRSA in ocular infections from India. The present study was carried out to study the changing trends of methicillin susceptibility in *S aureus* isolated from different ocular infections and to analyse the susceptibility of MRSA to quinolones and vancomycin.

We performed a retrospective review of microbiology records from January 2006 through December 2008 to determine the susceptibility pattern of *S aureus* isolated from corneal scrapings, vitreous, conjunctival swabs and lacrymal sac and adnexa to oxacillin (methicillin). *Staphylococcus* was identified to species level by conventional biochemical tests and by using Mini API (bioMérieux, France). Antibiotic susceptibility of the isolates was determined by using the Kirby Bauer disk diffusion method on cation-adjusted Mueller Hinton agar to oxacillin, ciprofloxacin, ofloxacin, moxifloxacin gatifloxacin and vancomycin. Plates were incubated at 35°C for 16–18 h in non–carbon dioxide incubator. The results were interpreted as per the Clinical Laboratory Standards Institute guideline.

STATISTICAL ANALYSIS

Changing the trends of oxacillin susceptibility in *S aureus* isolates was analysed by using χ^2 test, and the susceptibility evaluation of MRSA to different antibiotics was done by correlation analysis.

A total of 199 *S aureus* isolates were isolated during the study period. Of these, 74 were isolated from the corneal scrapings; 43, from the conjunctival swabs; 11, from the vitreous and 71 from the lacrimal sac and adnexa. Of 199 isolates, 68 (34.1%) were resistant to methicillin. Of 68 MRSA isolates, 22 (32.3%) were isolated from the conjunctiva, 23 (33.8%) were isolated from the lacrimal sac and adnexa, 20 (29.4%) were isolated from the corneal scrapings and 3 (4.4%) were isolated from the vitreous. All the 199 (100%) *S aureus* isolates were sensitive to vancomycin.

Yearwise distribution of *S aureus* isolates and their susceptibility to methicillin are shown in table 1.

The susceptibility rates of methicillin-resistant and methicillin-susceptible *S aureus* (MSSA) to quinolones are shown in table 2.

Methicillin-resistant *S aureus* causes severe ocular infections,[3 4] and prevalence of ocular MRSA infections varies from 3% to 30%.[1 5] In this laboratory-based investigation, the proportion of MRSA ocular infections increased from 26% in 2006 to 38% in 2008. An annual increase in MRSA incidence was observed in our study.

Hospital-acquired MRSA strains are usually multidrug resistant.[6] Community-acquired MRSA, though resistant to methicillin and other β lactam antibiotics (penicillin, cephalosporins and carbapenems), often remain sensitive to many other

Table 1 Susceptibility of *S aureus* to methicillin

Year	Methicillin sensitive (%)	Methicillin resistant (%)	95% CI	p Value
2006	33 (73.3)	12 (26.7)	13.3 to 38.7	0.075
2007	59 (64.9)	32 (35.1)	25.2 to 45.2	
2008	39 (61.9)	24 (38.1)	26.1 to 50.1	

Table 2 Susceptibility of methicillin-resistant and methicillin-susceptible *S aureus* to quinolones

	Cip				Mox				Gat				Of			
Met	S	R (%)	95% CI	p Value	S	R (%)	95% CI	p Value	S	R (%)	95% CI	p Value	S	R (%)	95% CI	p Value
S	64	67 (51)	42.5 to 59.7	<0.01	74	57 (43.5)	35 to 52	0.026	114	17 (13)	7.2 to 18.7	0.071	65	66	41 to 59.8	<0.01
R	20	48 (70)	59.7 to 81.41		27	41 (60)	48.6 to 71.9		52	16 (23)	13.4 to 33.6		17	51	64 to 856.9	

Cip, ciprofloxacin; Gat, gatifloxacin; Met, methicillin; Mox, moxifloxacin; Of, ofloxacin; R, resistant; S, susceptible.

CAREERS

Race, Disadvantage and Faculty Experiences in Academic Medicine

Linda Pololi, MBBS, MRCP[1], Lisa A. Cooper, MD, MPH[2], and Phyllis Carr, MD[3]

[1]Women's Studies Research Center, National Initiative on Gender, Culture and Leadership in Medicine: C - Change, Brandeis University, Waltham, MA, USA; [2]Johns Hopkins Medical Institutions, Baltimore, MD, USA; [3]Boston University School of Medicine, Boston, MA, USA.

BACKGROUND: Despite compelling reasons to draw on the contributions of under-represented minority (URM) faculty members, US medical schools lack these faculty, particularly in leadership and senior roles.

OBJECTIVE: The study's purpose was to document URM faculty perceptions and experience of the culture of academic medicine in the US and to raise awareness of obstacles to achieving the goal of having people of color in positions of leadership in academic medicine.

DESIGN: The authors conducted a qualitative interview study in 2006–2007 of faculty in five US medical schools chosen for their diverse regional and organizational attributes.

PARTICIPANTS: Using purposeful sampling of medical faculty, 96 faculty were interviewed from four different career stages (early, plateaued, leaders and left academic medicine) and diverse specialties with an oversampling of URM faculty.

APPROACH: We identified patterns and themes emergent in the coded data. Analysis was inductive and data driven.

RESULTS: Predominant themes underscored during analyses regarding the experience of URM faculty were: difficulty of cross-cultural relationships; isolation and feeling invisible; lack of mentoring, role models and social capital; disrespect, overt and covert bias/discrimination; different performance expectations related to race/ethnicity; devaluing of research on community health care and health disparities; the unfair burden of being identified with affirmative action and responsibility for diversity efforts; leadership's role in diversity goals; and financial hardship.

CONCLUSIONS: Achieving an inclusive culture for diverse medical school faculty would help meet the mission of academic medicine to train a physician and research workforce that meets the disparate needs of our multicultural society. Medical school leaders need to value the inclusion of URM faculty. Failure to fully engage the skills and insights of URM faculty impairs our ability to provide the best science, education or medical care.

Received January 29, 2010
Revised May 26, 2010
Accepted July 22, 2010
Published online August 10, 2010

KEY WORDS: medical faculty; underrepresented minorities; race.
J Gen Intern Med 25(12):1363–9
DOI: 10.1007/s11606-010-1478-7

INTRODUCTION

Medical schools hold a social mission to educate physicians who will care for the entire population.[1,2] Diversity among faculty enhances the ability of academic medicine to fulfill its educational, research and patient-care missions.[3] Inclusion of under-represented minority faculty members (URM) in medical schools promotes more effective health care delivery to a diverse population; improves the quality of medical education,[4] and may stimulate research attentive to the needs and concerns of minority groups.[5]

Despite these compelling reasons to draw on the perspectives and contributions of URM faculty members, there is an alarming dearth of these faculty in US medical schools and a serious paucity in leadership or senior roles[6–8] (Table 1).

Studies have shown that URM faculty are less satisfied and more likely to leave academic medicine, advance more slowly and are less likely to be in the basic sciences.[9–12,17,18] Additionally, minority faculty report experiences of ethnic harassment, biased treatment and racial "fatigue."[11,13–16] They spend more time in patient care and less in research than their non-minority colleagues. Efforts have increased the enrollment of URM medical students,[19–22] but the environment or culture for URM faculty has received much less attention.

Out of concern regarding the failure of academic medicine to adequately recruit, retain and advance diverse faculty, we formed a national collaborative, the National Initiative on Gender, Culture and Leadership in Medicine ("C - Change")[23] to engage five US medical schools in action research to facilitate culture change in academic medicine. In this partnership, we conducted an in-depth interview study of faculty to deepen our understanding of factors underlying the lack of URM faculty in the nation's medical schools.

METHODS

As part of the larger C - Change initiative, we selected five schools representing organizational characteristics of all US medical schools, [i.e., public (two)/private (three), NIH research intensive (two), primary care orientation/community orienta-

Table 1. Representation of Faculty Members of African-American/Hispanic/Latino and Native American Groups is Far Below the Demographics of These Groups In the US Population and US Medical Students.[6–8]

	US population 2000[6]	Medical students 2007[7]	All medical faculty 2007[7]	Medical faculty instructors/ assistant professors 2007[8]	Medical faculty associate/ full professors 2007[8]
African American	12.3%	6.4%	3.0%	8.0%	3.8%
Hispanic/Latino	12.5%	7.2%	4.2%	3.0%	2.2%
Native American	0.9%	0.3%	<0.1%		

tion (one)]. The sample represented all designated Association of American Medical Colleges (AAMC) regions. Sex and URM faculty demographics in these five schools were almost identical to national statistics.

Participant Criteria

Using purposeful and chain referral sampling,[24] we interviewed equal numbers of faculty from each school stratified by sex, race/ethnicity, department/discipline and career stage. Participants were research scientists, medical and surgical subspecialists, and generalist faculty holding doctoral degrees (84% MD/MBBS, 16% PhD) and represented 26 disciplines. The 96 faculty members interviewed in 2006–2007 represented (almost equally) four career stages: (1) early career (initial faculty appointment for 2 to 5 years), (2) "plateaued," (faculty for ≥10 years and who had not advanced as expected), (3) faculty in leadership roles such as deans, department chairs and center directors, and (4) former faculty who had left academic medicine. We oversampled women (55%) and URM. The 17% African American/Black, 4% Hispanic/Latino respondents represented 11 specialties/disciplines.

Data Collection and Analysis

A multidisciplinary research team (2 MDs, 2 PhDs) conducted in-depth, semistructured interviews—15% in person and 85% by telephone for convenience. Interviews (typically 1 h) were audio-recorded and transcribed verbatim. The interview guide (developed from pilot interviews) consisted of open-ended questions on aspirations of faculty, energizing aspects of their careers, barriers to advancement, collaboration, leadership, power, values alignment and work-family integration (Table 2). We concluded interviewing when we no longer obtained new information.

After deletion of all identifying information, transcribed data were coded and organized using Atlas.ti software. Inductive analysis [25,26] identified patterns and themes as they emerged from the coded data. The secondary analysis reported here utilized all coded data related to URM faculty and discrimination, and the entire interviews of URM faculty. We verified our findings using a consensus process. Brandeis Institutional Review Board approved the project. The example quotations illustrate themes stated repeatedly.

RESULTS

Male and female URM faculty experienced: difficulty in cross-cultural communication; feeling isolated and invisible; lack of mentoring, role models and social capital; disrespect, overt and covert bias and racism; devaluing of professional interests; being identified with affirmative action programs and diversity responsibilities; and financial hardship.

Cross-Cultural Communication

Many URM faculty described problems with conversation and relationship formation with Caucasian colleagues, e.g., *"Maybe they don't know how to talk to me because I'm an African American person."* They ascribed this to having different professional and social frames of reference. URM faculty didn't feel included and perceived that they caused Caucasians to be uncomfortable in conversing with them. This lack of connection created a barrier to collaborating with other faculty.

> It makes me feel like they're so uncomfortable. We don't have the same frames of reference. And it doesn't feel comfortable on either side of the conversation. I feel like I'm making people think about things they don't want to think about and so why bother? (URM female, plateaued)

Some described academic medicine as feeling like a foreign culture.

> So academic medicine is a foreign culture that isn't friendly to American Indians and Latinos. You're not going to attract Latinos, American Indians who have a community bent, who want to change social systems, who have a sense of family and community. It's very hard for us to fit in academic institutions, where that's about the individual. (URM male, early career)

Table 2. Interview Guide Questions

What is it about your work that energizes you?
When have you felt most successful in your work?
What's been your sense of being a part of your institution?
What has been difficult or frustrating in your work?
Can you talk about some experiences you've had related to the advancement of your own career in academic medicine?
What do you see as valued at your institution?
What does it take to get into a position of power or leadership in your school?
Is leadership something you've wanted or want for yourself?
How has power affected you/your career in academic medicine?
How do your personal values align or conflict with what you experience in academic medicine?
Why do you think there are so few women in the upper reaches of academic medicine?
What about underrepresented ethnic groups or people of color?
Tell me about the relationship between your work and family or personal life.
How are your aspirations for yourself in academic medicine being fulfilled?

Isolation

Isolation due to the scarcity of colleagues of color was cited frequently. This was especially prominent for female faculty.

> Early on, there were just no women and certainly, no faculty of color, and so you're just there by yourself. (URM female, leader)

Some URM faculty remarked that people of color need other supportive relationships from family, church or community outside medicine to survive the professional isolation.

Feeling Invisible

The sense of isolation was compounded by feeling invisible in the institution and at national professional meetings.

> What I struggled with for a long time here was my being an African-American woman, in a male, white male-dominated institution and the feeling that I was invisible. My opinion didn't matter, what I was feeling didn't matter. There were people who I passed every single day, who were chairmen of departments, and I mean, good God, after 5 years you've got to see a person… I would really hate to go to national meetings. Because I wouldn't see very many people who looked like me and even though I had met people the year before or 2 years before, they always acted as if they'd never seen me before. (URM female, plateaued)

Lack of Mentoring, Role Models and Social Capital

Many commented on the lack of support and mentoring and the paucity of URM role models in academic medicine. This was especially significant as many African American and Hispanic faculty came from backgrounds where they had little exposure to academia or the systems of higher education.

> When you're the first person in your family to reach this point, you are clueless. I was not receiving any counseling at all about what the next move was… A lot of people of color don't know that. (URM female, leader)

Programs for URM students and resident were available, but faculty believed it "*pretty well peters out at the faculty level*." Many faculty of color acknowledged that they lacked role models for themselves, but still felt responsible for serving in leadership roles and being role models for other URM faculty.

> The other reason that leadership is important for me is because you want your family, other people of color to say, "Okay, I can do this too, I can do it." Academics is not an area that people think about. (URM female, leader)

> I'm at a point in my career where I have to decide whether to stay or go, and if I go, where's the role model for the ones coming behind me? Then they have a similar experience like I did. (URM female, plateaued)

Disrespect, Prejudice, Bias and Racism

We heard numerous accounts of experiencing racist remarks and bias. Minority faculty described being stereotyped; sometimes being viewed as similar to uneducated minority patients or other people of color in service roles.

> If the majority of the patients that you're treating are African American and very poor and uneducated, and I'm African American, well, people are sometimes not able to make the distinction between some of those patients and you. (URM female, early career)

> So I showed up at the meeting last year, and one of the Division Directors asked me to take his luggage to his room…I was just puzzled. And he said, "You are at this meeting aren't you?" And then it became clear to me that he thought I was one of the organizers of the meeting. I said, "Well, yes I am at the meeting." And then he very sharply said, "Well then can you take care of this?" And I said: "Sir, I believe I'm at the meeting for the same reason you are." (URM female, leader)

Another minority physician remembered an incident as an intern.

> I was on call and one of the nurses interrupted me and said, "Oh go to room such and such, the sheets need to be changed." …making the assumption that if I am African-American, I'm here to clean the beds. (URM female, left academic medicine)

A pervasive example of stereotyping was that colleagues and supervisors often had low or mediocre expectations of what URM faculty (or students) could accomplish.

> I have heard it from African-American students that were very interested in science, and they had teachers that said, "I don't think you really can get a PhD." (URM female early-career)

Individually, URM faculty consistently believed that they had to perform at a higher level than others in similar situations to be perceived as accomplished.

> I think you always feel like you're expected to do a mediocre job. Always. And so, you strive to be super woman. To combat the expectation that you're only going to be mediocre. (URM female, plateaued)

Others recounted instances of racism.

> I sent my resume for something and when I showed up someone said to me, "Your resume didn't look black." Can you imagine someone saying that? (URM female, senior)

> There was a night shift that I worked, the resident came down and asked me if we could hold a patient in the emergency department because it had been a busy night for the resident, and he didn't want to admit the patient. And I said "No," because the patient was an older woman on a stretcher down here in the ED, I

> wanted her to go upstairs. He walked away and mumbled, "You black bitch." My boss happened to come down first thing in the morning, and I recounted the episode to him, and told him I was so angry I could have punched this guy and he says to me, "Well, you know, we don't live in the jungle." That was his response. I will never forget that. (URM female, plateaued)

Discrimination in recruitment emerged as a sub-theme. One respondent recounted her experience of discrimination by other faculty members. She pointed out the burden of dealing with discrimination for many years.

> So he asked the chair to bring me on as a faculty member. And one faculty went to the Dean, and said: "Let's not bring her on. Let's wait 'till next year because we want (name), because we believe that a white, Jewish male will fit the environment better." So, sometimes you look back and wonder why you stayed. (URM female, leader)

Another Caucasian faculty explained discrimination on the basis of class. He believed that prejudice exists against non-white speech patterns.

> I think that even though people may not be prejudiced against skin color, they may be prejudiced against—and I know I'm prejudiced—language patterns. It's the way I'm prejudiced against people who put their Rs in the wrong place. It's a sign of class. (Non-minority male, plateaued)

Devaluing of Professional Interests

In the departments where there was focus on research on underserved populations or community-based health care, URM faculty perceived a more favorable microenvironment for themselves. Such departments were better able to recruit URM faculty. Conversely, respondents commented on how research in communities and giving back to their communities, as well as research on minority health disparities (MHD) was less rewarded, and perceived as less weighty.

> I think from all the other departments' perspective they'd say, "What's going on with them? They're doing soft research out in the community and they're not real scientists." (URM male, left academic medicine)

Minority faculty felt a sense of responsibility to their communities that often conflicted with the demands of an academic career.

> It's important that we're out in the community actually caring for patients, giving back to our community. I know that my closest colleagues who are African-American definitely feel that way. They felt that they couldn't do that within the confines of academic medicine. (Non-minority male, left academic medicine)

Additionally, issues of tokenism and "window dressing" were voiced. Some faculty suggested that URM faculty are "*just a pawn to be used by the institution*" to show that the institution is attending to URM recruitment or doing something about MHD.

> The minorities see academia as an environment that they can potentially thrive in and they get played. They become the representative in that it shows that the institution is doing something about health disparities and they get used, and they also don't wind up in the decision-making circles. (Non-minority male, plateaued)

Burden of Affirmative Action Programs and Representing One's Own Race

Compounding tokenism, faculty spoke of the stigmata of having participated in affirmative action programs or programs specifically targeting MHD, thus being simultaneously benefited and disadvantaged. URM perceived that others thought they had *"got there because they were Black."*

> You almost feel that you have to do better than anybody else to prove that you are where you are because you deserve it. I was very lucky and I got my R01 very quickly and I got a very, very good score. I was really proud of that. I worked very hard on that grant. And a colleague of mine, he looked at me and said, "I'm convinced that these things are decided based on ethnicity." (URM female, leader)

Often URM faculty were asked to provide service and committee work to promote diversity. They felt conflicted in the realization that this service on behalf of URM detracted from personal scholarship and an individual need to advance in the system.

> They told me the only reason I got the job was because I was Black. And when I came into the Dean's office, there was talk of: "We don't have an Office of Minority Affairs (OMA) here." Every time they start talking about an OMA they start looking at me. And I said, "If I'm good enough to be the Dean of Minority Students, I'm good enough to be the Dean of all the students." (URM female, plateaued)

A dilemma for many faculty was how to manage concurrent efforts to take care of other people of color, as well as to advance professionally to be accomplished role models.

Responsibilities of Leadership

Many faculty commented on the pivotal role of leaders with respect to diversity and the scarcity of people of color in leadership roles in medical schools. Leaders were perceived as not valuing diversity and needing to make a firm commitment to diversity goals if to acheive real change. Leaders tolerated unacceptable behavior or even racism (e.g., see Racism section). Interviewees believed that leaders rarely selected people of color for leadership positions as doing so would detract from their sense of comfort interacting with people like themselves.

> It has to be something that leaders prioritize. When everything else is equal, you have to step up to the plate

and choose the person who does bring diversity. So you need leaders who are willing to stand for it. (URM male, senior)

Some commented that having more URM faculty in positions of leadership would dispel stereotypic myths and create greater exposure to minority excellence on an individual basis.

An intern on my teaching service said to me "I really enjoyed working with you. I truly respect the way that you take care of patients and I want to try to mimic the way I take care of patients after you." He was a non-minority young man and I thought that was incredibly important for me to be in a position to have somebody like him say that. But you have to have the commitment from schools, from hospitals, from administrators to find the people to be in those [leadership] positions and I think it has a huge calming effect on society in general. (URM male, early career)

URM faculty reported that they usually have to be the one to notice and comment on inequity and that this responsibility is not assumed by majority leaders. Some commented that when URM faculty assume administrative leadership, it's at the expense of advancing their own scholarship. On the other hand, having an URM in a leadership role gives the clear message about commitment to achieving diversity.

Some drew a contrast between African-American, and Asian or Middle Eastern faculty. perceiving that the latter groups often came from educated and privileged families who were more familiar with academic pursuits and hierarchies:

I think Asian and Middle Eastern men have been accepted much more than African-American men, and that reflects who's going to medical school. Second and third generation people from Pakistani, Indian and Iranian families. They're well trained and they're very hard workers, and excellent clinicians and teachers—so no issues there. But I think that gives the institution the feeling that they're ethnically diverse, but with all due respect, these are "WASPs" with brown skin. They're more similar in their behavior to the white Anglo-Saxon Protestant model than a Brooklyn Jew, for instance, who's noisy and loud. (Non-minority female, leader)

This white woman shows a nascent understanding of a certain way one has to act in order to be accepted, i.e., similarly to a white male. Asian and Middle-Eastern faculty with more privileged backgrounds may more easily adapt to expected behavior patterns and consequently advance more frequently than URM faculty.

Financial Hardship

A persistent interview theme was the financial sacrifice perceived by choosing a career in academic medicine.

So they are seen as not just the breadwinners for their household, but for the greater family at large. There's a sense of responsibility not only to give back to the community, but also to earn a higher wage to help out the extended family. (Non-minority male, left academic medicine)

Slowness of Change

The history of segregation and slavery in the US is still a part of many people's consciousness. Several faculty highlighted the slowness of realizing the full positive outcomes of legislation resulting from the civil rights movement. *"I'm just disappointed with the progress of our country."*

I mean it's your normal change process, establishing the value of the differing person or persons and eventually change occurs. It's the same process that we went through with integration. That wasn't overnight either. And we're still struggling with it in academic medicine. (Non-minority male, senior)

DISCUSSION

URM faculty bring knowledge and experience of different backgrounds and world views to medical schools. Our findings suggest that these valuable attributes and abilities, instead of being perceived and received as beneficial, are often responded to as untoward contributions and become barriers to acceptance in the systems of academic medicine.

Isolation and feeling like an outsider resulted from a combination of barriers to communication and relationship formation with majority faculty; scarcity of faculty of color; and lack of role models. Lack of family instrumental support and social capital combined with education-related debt added to the burden of trying to advance professionally. Faculty experienced disrespect, discrimination, racism and a devaluing of their professional interests in community service and MHD. Women faculty commented on the double disadvantage of gender and minority status.

URM faculty experience social as well as professional discrimination and may feel justifiably angry. The "tokenism" and "window dressing" they describe pertains to at least three concepts: a lack of authenticity among institutional leaders in efforts to include minorities; the burden of having to represent one's entire race; and being on the receiving end of special programs and assumptions that the achievements of people of color are due to special favors rather than merit. Faculty ascribe a pivotal role to leadership in combating discrimination and achieving diverse faculties. Many leaders lack the experience of having different types of people in leadership roles, and it may seem risky to put power in the hands of less experienced people. Most URM faculty come from non-affluent families (in contrast to many white majority students) and incur substantial debt during training. URM physicians supported their households and often their extended families. The combination of this and dedication to their communities contributed to URM leaving academic medicine.

While published research on diversity in medical school faculty report on a single school,[18] on URM physicians in practice[15] and some national recommendations,[27] this paper's contribution is in-depth data on the experience of URM medical faculty from diverse subspecialties, collected from five disparate schools in different US regions. While we have

focused on URM faculty, other faculty of color may contend with many of the same disadvantages. The findings on relational barriers align with our study results in non-minority faculty.[28] Limitations of the study are those inherent in qualitative studies with relatively small numbers of participants: selection or sampling bias, potential for response bias and the subjective nature of analytic strategy. Even so, qualitative studies singularly allow voicing of perceptions of individuals who voluntarily share such information. We found the themes to be consistent and highly congruent for faculty of varying rank, discipline and sex across the five schools.

McIntosh observed that people who benefit most (in the short term) from privilege systems are mostly unaware of and blind to the existence of privilege systems. This preserves the myths of moral and managerial meritocracy.[29] This likely occurs because the exposure of bias is often painful and disturbing, particularly among individuals who explicitly hold egalitarian and humanitarian views. Having inherited unconscious biases that are manifested unintentionally in interpersonal interactions, these individuals may feel guilty about their own advantage (acquired typically without effort or consent on their part) and its role in keeping others disadvantaged. Through elucidation of URM faculty experiences, we hope to raise awareness among health professionals, educators, administrators and policy-makers of obstacles to achieving the goal of having URM faculty as leaders in academic medicine.

Medical schools and their policies need to reward service and research on community-based health care and MHD, similarly to other accomplishments and research if this work is to be shouldered by a broader set of faculty. Health disparities in the US are among the highest in the developed world, and reducing them is a major health priority.[30–32] Successful strategies to reduce disparities must address the physician workforce.[27,33] We propose that having more URM faculty in senior and leadership roles in medical schools will support training a more diverse physician population and increase the cultural awareness and skills of all physicians-in-training and biomedical scientists. These factors will contribute to a greater capacity to care for underserved groups and to better elucidate the causes of and solutions to health disparities. Failure to fully engage the skills and insights of URM faculty means that we don't have the best science and the best medical care that we could have. We agree that medical schools and their leadership should be evaluated on the extent to which their graduates meet the health needs of the nation.[33–35] Achieving a diverse medical school faculty would help meet the institutional mission of academic medicine to train a physician and research workforce that meets the needs of our multicultural society.

ACKNOWLEDGMENTS: *The authors gratefully acknowledge the critical funding support of the Josiah Macy, Jr., Foundation and the supplemental funds to support data analysis provided by the US Office of Public Health and Science, Office on Women's Health and Office on Minority Health; the National Institutes of Health, Office of Research on Women's Health; the Agency for Healthcare Research and Quality; the Centers of Disease Control and Prevention, and the Health Resources and Services Administration (contract: HHSP233200700556P). The authors thank Peter Conrad and Sharon Knight, who participated in data collection and data coding, and Kerri O'Connor for manuscript preparation. The authors are indebted to the medical faculty who generously shared their experiences in the interviews.*

Contributors: *Specific Contributions From Each Author*

Pololi: conception, design, data collection, analysis and interpretation, drafting the article, final approval.

Cooper: analysis and interpretation, drafting the article, final approval.

Carr: data collection and coding, manuscript review, final approval.

Funders: *United States Office of Public Health and Science, Office on Women's Health, Contract: HHSP233200700556P*

Josiah Macy, Jr. Foundation

Prior Presentations: *Society of General Internal Medicine, 33rd Annual Meeting, 2010. Research plenary presentation.*

Conflict of Interest: *None disclosed.*

Ethical Approval: *Brandeis University Institutional Review Board for the protection of human subjects approved the study.*

Corresponding Author: *Linda Pololi, MBBS, MRCP; Women's Studies Research Center, National Initiative on Gender, Culture and Leadership in Medicine: C - Change, Brandeis University, Mailstop 088, Waltham, MA 02454-9110, USA (e-mail: lpololi@brandeis.edu; URL: http://cchange.brandeis.edu).*

REFERENCES

1. **McCurdy LL, Goode D, Inui TS, Daugherty RM Jr, Wilson DE, Wallace AG, Weinstein BM, Copeland EM 3rd.** Fulfilling the social contract between medical schools and the public. Acad Med. 1997;72(12):1063–70.
2. **Nickens H, Smedley B.** The right thing to do, the smart thing to do: enhancing diversity in the health professions: summary of the symposium on diversity in health professions in honor of Herbert W. Nickens. Washington: National Academy Press: Institute of Medicine; 2001.
3. **Kington R, Tisnado D, Carlisle D.** Increasing the racial and ethnic diversity among physicians: an intervention to address health disparities. In: **Smedley BD, Colburn L, Evans CH, eds.** The right thing to do, the smart thing to do: enhancing diversity in the health professions. Washington: National Academies Press; 2001:64–8.
4. **Whitla DK, Orfield G, Silen W, Teperow C, Howard C, Reede J.** Educational benefits of diversity in medical school: A survey of students. Acad Med. 2003;78(5):460–6.
5. **Collins KS, Hughes DL, Doty MM, Ives BL, Edwards JN, Tenney K.** Diverse communities, common concerns: Assessing health care quality for minority Americans. Findings from the Commonwealth Fund 2001 Health Care Quality Survey. New York: The Commonwealth Fund; 2002.
6. Census 2000: United States Profile. US Census Bureau web site. Available at: http://www.census.gov/prod/2002pubs/c2kprof00-us.pdf. Accessed: July13, 2010.
7. **Castillo-Page L.** Diversity in medical education: Facts & figures 2008. Washington: Association of American Medical Colleges; 2008.
8. Association of American Medical Colleges. Faculty Roster 2008. Available at: http://www.aamc.org/data/facultyroster/. Accessed: July 13, 2010.
9. **Palepu A, Carr PL, Friedman RH, Amos H, Ash AS, Moskowitz MA.** Minority faculty and academic rank in medicine. JAMA. 1998;280:767–71.
10. **Palepu A, Carr PL, Friedman RH, Ash AS, Moskowitz MA.** Specialty choices, compensation, and career satisfaction of underrepresented minority faculty in academic medicine. Acad Med. 2000;75:157–60.
11. **Peterson NB, Friedman RH, Ash AS, Franco S, Carr PL.** Faculty self-reported experience with racial and ethnic discrimination in academic medicine. J Gen Intern Med. 2004;19(3):259–65.
12. **Smedley BD, Stith AY, Colburn L, Evans CH.** The right thing to do, the smart thing to do enhancing diversity in health professions. Institute of Medicine. Washington: National Academy Press; 2001.

13. **Corbie-Smith G, Thomas TB, Williams MV, Moody-Ayers S.** Attitudes and beliefs of African Americans toward participation in medical research. J Gen Intern Med. 1999;14:537–46.
14. **Price EG, Gozu A, Kern DE, Powe NR, Wand GS, Golden S, Cooper LA.** The role of cultural diversity climate in recruitment, promotion, and retention of faculty in academic medicine. J Gen Intern Med. 2005;20 (7):565–71.
15. **Nunez-Smith M, Curry L, Bigby J, Berg D, Krumholz HM, Bradley EH.** Impact of race on the professional lives of physicians of African descent. Ann Intern Med. 2007;146(1):45–51.
16. **Carr PL, Palepu A, Szalacha L, Caswell C, Inui T.** Flying below the radar: a qualitative study of minority experience and management of discrimination in academic medicine. Med Edu. 2007;41(6):601–9.
17. **Fang D, Moy E, Colburn L, Hurley J.** Racial and ethnic disparities in faculty promotion in academic medicine. JAMA. 2000;284(9):1085–92.
18. **Price EG, Powe NR, Kern DE, Golden SH, Wand GS, Cooper LA.** Improving diversity climate in academic medicine: faculty perceptions as a catalyst for institutional change. Acad Med. 2009;84:95–105.
19. **Aagaard EM, Julian K, Dedier J, Soloman I, Tillisch J, Pérez-Stable EJ.** Factors affecting medical students' selection of an internal medicine residency program. J Natl Med Assoc. 2005;97(9):1264–70.
20. **Dyrbye LN, Thomas MR, Huschka MM, Lawson KL, Novotny PJ, Sloan JA, Shanafelt TD.** A multicenter study of burnout, depression, and quality of life in minority and nonminority US medical students. Mayo Clin Proc. 2006;81(11):1435–42.
21. **Odom KL, Morgan Roberts L, Johnson RL, Cooper LA.** Exploring obstacles to and opportunities for professional success among ethnic minority medical students. Acad Med. 2007;82(2):146–53.
22. Association of American Medical Colleges. The Diversity Research Forum: Successfully evaluating diversity efforts in medical education. Washington: AAMC; 2007.
23. National Initiative on Gender, Culture and Leadership in Medicine: C - Change website. Available at: http://cchange.brandeis.edu. Accessed: July 13, 2010.
24. **Biernacki P, Waldorf D.** Snowball sampling: problems and techniques of chain referral sampling. Sociol Methods Res. 1981;10:141–63.
25. **Glaser BG, Strauss AL.** The discovery of grounded theory: strategies for qualitative research. Aldine Transaction: Chicago; 1967.
26. **Charmaz K.** Constructing grounded theory: a practical guide through qualitative analysis. Thousand Oaks (CA): Sage Publications; 2006.
27. **The Sullivan Commission.** Missing persons: Minorities in the health professions diversity. Washington (DC): The Sullivan Commission; 2004.
28. **Pololi L, Conrad P, Knight S, Carr P.** A study of the relational aspects of the culture of academic medicine. Acad Med. 2009;84:106–14.
29. **McIntosh P.** White privilege, color and crime: A personal account. In: **Mann CR, Zatz MS, eds.** Images of color, images of crime. Los Angeles (CA): Roxbury Publishing Company; 1998.
30. AHRQ (2004). National Healthcare Disparities Report: Department of Health and Human Services, US Department of Health and Human Services.
31. **Kelley E, Moy E, Stryer D, Burstin H, Clancy C.** The national healthcare quality and disparities reports: an overview. Med Care. 2005;43(3):I3–8.
32. **Siegel S, Moy E, Burstin H.** Assessing the nation's progress toward elimination of disparities in health care. J Gen Intern Med. 2004;19 (2):195–200.
33. **Freeman J, Ferrer R, Greiner A.** Viewpoint: Developing a physician workforce for America's disadvantaged. Acad Med. 2007;82(2):133–8.
34. **Mullen F, Chen C, Pettersons S, Kolsky G, Spagnola M.** The social mission of medical education: Ranking the schools. Ann Intern Med. 2010;152:804–11.
35. **Pololi L, Kern DE, Carr P, Conrad P.** Authors' Reply: Faculty Values. J Gen Intern Med. 2010;25(7):647.

Choice of Initial Antiepileptic Drug for Older Veterans: Possible Pharmacokinetic Drug Interactions with Existing Medications

Mary Jo V. Pugh, PhD,[ab] Anne C. VanCott, MD,[cd] Michael A. Steinman, MD,[ef] Eric M. Mortensen, MD,[ag] Megan E. Amuan, MPH,[h] Chen-Pin Wang, PhD,[ab] Janice E. Knoefel, MD,[ijf] and Dan R. Berlowitz, MD, MPH[hl]

OBJECTIVES: To identify clinically meaningful potential drug–drug interactions (PDIs) with antiepileptic drugs (AEDs), the AEDs and co-administered drugs commonly associated with AED-PDIs, and characteristics of patients with high likelihood of AED-PDI exposure.

DESIGN: Five-year retrospective cohort study of veterans with new-onset epilepsy.

SETTING: National Veterans Affairs and Medicare databases.

PARTICIPANTS: Veterans aged 66 and older with a new diagnosis of epilepsy between October 1, 1999, and September 30, 2004 (N = 9,682).

MEASUREMENTS: AED-PDI was restricted to clinically meaningful PDIs identified using prior literature review. AED-PDIs were identified using participants' date of initial AED prescription and overlapping concomitant medications. Logistic regression analysis identified factors associated with AED-PDI, including demographic characteristics, chronic disease states, and diagnostic setting.

RESULTS: AED-PDI exposure was found in 45.5% (4,406/9,682); phenytoin, a drug with many PDIs, was the most commonly prescribed AED. Cardiovascular drugs, lipid-lowering medications, and psychotropic agents were the most commonly co-administered AED-PDI medications. Individuals with AED-PDI exposure were more likely to have hypertension (odds ratio (OR) = 1.46, 99% confidence interval (CI) = 1.24–1.82) and hypercholesterolemia (OR = 1.40, 99% CI = 1.24–1.57) than those without and to be diagnosed in an emergency or primary care setting than a neurology setting (emergency: OR = 1.30, 99% CI = 1.08–1.58; primary care: OR = 1.29 99% CI = 1.12–1.49).

CONCLUSION: Exposure to AED-PDI was substantial but less common in patients with epilepsy diagnosed in a neurology setting. Because potential outcomes associated with AED-PDI include stroke and myocardial infarction in a population already at high risk, clinicians should closely monitor blood pressure, coagulation, and lipid measures to minimize adverse effects of AED-PDIs. Interventions to reduce AED-PDIs may improve patient outcomes. **J Am Geriatr Soc 58:465–471, 2010.**

Key words: drug–drug interaction; epilepsy; geriatrics; antiepileptic drugs

Although many consider epilepsy to be a disease of childhood, the incidence of epilepsy is highest in older adults, with prevalence estimates between 1.8% and 2%.[1] Management of epilepsy in older patients is clinically challenging. In addition to concerns about ensuring that seizures be controlled with minimal side effects, older patients tend to have a number of comorbid conditions which require medications,[2] many of which have the potential to interact with commonly used antiepileptic drugs (AEDs).

A number of published reviews of treatment of older patients with epilepsy highlight, but fail to quantify, the risk of potential pharmacokinetic drug–drug interactions (PDIs) in individuals receiving older AEDs that are hepatically metabolized (e.g., phenytoin, carbamazepine, phenobarbital).[3–5] These drug–drug interactions may cause AED toxicity or reduce the efficacy of the interacting drug.[6,7] Because many of the drugs with potential interactions with

From the [a]South Texas Veterans Health Care System, VERDICT REAP, San Antonio, Texas; [b]Department of Epidemiology and Biostatistics and [g]Division of General Internal Medicine, Department of Medicine, University of Texas Health Science Center at San Antonio, San Antonio, Texas; [c]Veterans Affairs Pittsburgh Healthcare System, Neurology Division, Pittsburgh, Pennsylvania; [d]Department of Neurology, University of Pittsburgh, Pittsburgh, Pennsylvania; [e]San Francisco Veterans Affairs Medical Center, San Francisco, California; [f]Department of Geriatrics, University of California at San Francisco, San Francisco, California; [h]Center for Health Quality Outcomes and Economic Research, Bedford Veterans Affairs Hospital, Bedford, Massachusetts; [i]Geriatrics/Extended Care Section, New Mexico Veterans Healthcare System, Albuquerque, New Mexico; [j]Division of Gerontology, Department of Medicine, and [k]Department of Neurology, University of New Mexico, Albuquerque, New Mexico; and [l]School of Public Health, Boston University, Boston, Massachusetts.

Address correspondence to Mary Jo V. Pugh, PhD, South Texas Health Care System: Audie L. Murphy Division (VERDICT 11C6), 7400 Merton Minter Boulevard, San Antonio, TX 78229. E-mail: pughm@uthscsa.edu

DOI: 10.1111/j.1532-5415.2010.02732.x

0002-8614/10/$15.00

AEDs are used to treat hypertension, hyperlipidemia, and coagulation disorders, prolonged exposure to AEDs and potentially interacting drugs (PDIs) may be associated with adverse outcomes such as falls and fractures, stroke, myocardial infarction, and mortality.[8–10]

Recently, it was found that the majority of patients in all age groups received AEDs with the potential for pharmacokinetic drug–drug interactions and that many also received medications that may interact with those AEDs,[11] although it is unclear the extent to which AEDs with the potential for these drug interactions were prescribed concomitantly with potentially interacting drugs. Thus, little is known about the extent to which AED-PDIs occur or whether clinicians were cognizant of, and avoided, potential interactions.[11]

This study used national databases from the Veterans Health Administration, Department of Veterans Affairs (VA), to identify the extent to which older veterans with new-onset epilepsy had AED-PDI exposure. Although pharmacodynamic drug–drug interactions are also possible, this article focuses exclusively on pharmacokinetic drug–drug interactions. The duration of AED-PDI exposure, older patients at greatest risk of AED-PDI exposure, and those most likely to have the potentially interacting drug continued after starting AEDs were further identified.

METHODS

Data

After approval was received from institutional review boards at the University of Texas Health Science Center at San Antonio and the Bedford and Hines VA Hospitals, national VA inpatient, outpatient, pharmacy and Medicare data October 1, 1999, to September 30, 2004 (FY00-FY04), linked with an encrypted identifier to identify older veterans ($\geq$66) with new-onset epilepsy within the VA medical system were gathered. These data also allowed identification of the initial AED regimen received, concomitant medications received at the time of the first VA AED prescription, comorbid conditions, and demographic variables.

Population and Setting

A previously validated algorithm[12,13] was used to identify elderly individuals with new-onset epilepsy. Using an algorithm described previously,[13] older veterans with new-onset epilepsy were identified using the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM), code for epilepsy (345.XX) or convulsion (780.39) in VA inpatient, outpatient, and Medicare databases, and who also received an AED from the VA outpatient pharmacy within 1 year of the initial epilepsy diagnosis. Prior research found that, of those identified with epilepsy using this algorithm, 98% had documented epilepsy in the electronic medical record.[13]

Measures

Dependent Variable: PDIs

PDIs were identified based on "clinically important drug interactions" based on pharmacokinetic properties described previously.[6] Clinically important drug interactions include PDIs that may affect the clinical management of the patient or that might have severe adverse outcomes for patients.[7] The assessment was further restricted to drugs for chronic conditions such as hypertension, hypercholesterolemia, and mental health conditions rather than acute conditions such as infection. Although there is a potential for interactions between AEDs, only a small number individuals received potentially interacting AEDs, and they also had AED-PDIs. Thus, the focus was exclusively on the matter of AED-PDI in this study. As such, these AED-PDIs do not completely reflect the VA drug–drug interaction screening criteria, which are more broadly defined and include drug interactions defined as mild[14] and drug interactions that are not relevant to older patients (e.g., oral contraceptives). Table 1 provides a list of the most common PDIs that are particularly relevant for older individuals and the potential effect of each AED-PDI. Most AED-PDIs are associated with older AEDs that are cytochrome P450 inducers (phenytoin, carbamazepine, phenobarbital). Moreover, Micromedex defines many of these AED-PDIs as moderate.[14]

The initial AED regimens prescribed within the VA were first identified. All AEDs prescribed the day of the initial VA prescription were considered in the assessment of PDI. Next all unique medications in the current drug regimen the day of the AED prescription were identified. These drugs were determined using an algorithm that identified the initial fill date plus the day supply of that fill and all subsequent fills plus 30 days, because most patients are not entirely adherent with medications and have some medications left over at the end of the fill period. For the initial approach, those receiving drugs for chronic conditions identified as having clinically meaningful drug interactions with the initial AED were classified as having AED-PDIs.[6] Because whether interacting drug prescriptions were discontinued could not be accurately determined, the final measure of AED-PDI was restricted to those who received a subsequent prescription of potentially interacting drugs after the initial AED prescription, which accounts for changes physicians made to drug regimens that may address the PDI. To further assess PDI exposure, the median duration of each AED-PDI for the year after the initial AED prescription was also determined. The day supply of medications during the year was calculated as long as the patient received both the AED and the potentially interacting drug.

Independent Variables

Demographic Characteristics

Using VA and Medicare data sources, the age, sex, and race of each patient were identified. VA data were also used to identify marital status (married vs not married). Although information on race is frequently missing from VA data, Medicare data were supplemented when VA data were missing for race, leaving less than 1% missing race data.

Diagnosis-Related Issues

Because prescribing patterns changed over the years of the study, with greater use of newer AEDs over time, the fiscal year of first AED prescription (FY00, FY01, FY02, FY03, FY04) was included in logistic regression analyses.[13] Furthermore, because individuals in this cohort are generally eligible to receive health care through Medicare, whether

Table 1. Clinically Meaningful Drug Interactions

Drug Class or Drug	Anticonvulsant	Interaction or Potential Outcomes
Cardiovascular drugs		
Disopyramide, mexiletine, and quinidine	CBZ, PHT, PB	Decreased antiarrhythmic concentrations
Propranolol, metoprolol	CBZ, PHT, PB	May require increase dosage of beta-blocker
Nifedipine, felodipine, and nimodipine	CBZ, PHT, PB	May nullify effects of the calcium channel blocker
Diltiazem	CBZ, PHT	Increase AED plasma concentration
Nimodipine	VPA	Increase plasma concentrations of nimodipine
Verapamil	CBZ	CBZ toxicity
Atorvastatin, fluvastatin, lovastatin, and simvastatin	CBZ, PHT, PB	May reduce statin efficacy
Ticlopidine	PHT, CBZ	CNS toxicity by elevated AED plasma concentration
Hematological agents		
Warfarin	CBZ, PHT	Decreased anticoagulant effect
Central nervous system agents		
Amitriptyline	CBZ, VPA	Increases metabolism of and reduces plasma concentration of amitriptyline
Bupropion, paroxetine	PHT, PB	Reduced plasma concentrations of the antidepressant
Nefazodone	CBZ	Contraindicated; CBZ toxicity and reduced effectiveness of nefazodone
Nortriptyline, clomipramine, and amitriptyline	PHT	Concurrent use can result in PHT toxicity
Nortriptyline, clomipramine	VPA	May inhibit metabolism of antidepressant causing elevated plasma concentrations
Chlorpromazine, clozapine, haloperidol, ziprasidone, olanzapine, quetipine, mesoridazine, risperidone	CBZ, PHT	Decreased antipsychotic plasma concentrations can result in therapeutic failure
Gastrointestinal agents		
Cimetidine	CBZ, PHT	AED toxicity
Sucralfate	PHT	Decreased phenytoin effectiveness
Omeprazole	CBZ, PHT	AED toxicity
Systemic anti-infective agents		
Erythromycin	CBZ, VPA	AED toxicity
Doxycycline	CBZ, PHT, PB	May reduce doxycycline effectiveness

Source: Patsalos and Perucca (2003).
AED = antiepileptic drug; CBZ = carbamazepine; PB = phenobarbital; PHT = phenytoin; VPA = valproic acid.

the epilepsy diagnosis was first documented in VA or Medicare data was also determined because treatment may have been initiated outside of the VA setting.

Comorbid Conditions

Validated ICD-9-CM code algorithms were used to identify comorbid conditions the patient experienced before the date of the initial epilepsy diagnosis.[15–18] Comorbidities identified included those that are commonly treated with drugs that interact with commonly used AEDs (e.g., hypertension, cardiovascular disease, diabetes mellitus, hypercholesterolemia, psychological disorders (any vs no psychological disorders)) and those comorbidities that have been associated with epilepsy in epidemiological studies (e.g., stroke, dementia, brain tumors, head injury).[19–22]

Medication Burden

The number of unique medications for each individual the year before inclusion in the cohort was tracked and considered to be a proxy for overall medication burden.[23] Initially, medication burden was classified according to quartiles to ease interpretation of data. Because odds ratios were the same for those in the second through fourth quartiles, those groups were combined to create a dichotomous measure: low medication burden (1–5 medications) versus high medication burden ($\geq$6 medications).

Analysis

Frequencies of demographic and clinical characteristics were calculated for the cohort, and the rate and duration of AED-PDI was identified overall and according to AED. Multivariable logistic regression analysis was then conducted to identify factors that were associated with AED-PDI and to identify characteristics of those for whom the offending drug was not renewed after initiation of the AED. Because of the large cohort size, $P < .01$ was used as the level of significance. Interactions were included in the initial models; because none were statistically significant, they were not included in the final model.

RESULTS

Demographic Characteristics

The cohort consisted of 9,682 older veterans who met the criteria for new-onset epilepsy. The cohort were mostly male (98.0%) and white (78.5%); representation of blacks (15.8%) and Hispanics (4.9%; 0.8% other (Native American, Pacific Islander, Asian)) was similar to that of the overall elderly population.[24] Only 4.6% were aged 85 and older; 48.1% were aged 75 to 84, and 47.3% were aged 66 to 74. The majority of the cohort (64.7%) was married.

Context of Epilepsy Diagnosis

Table 2 shows that the incidence of epilepsy was relatively consistent over the study period and that 61.0% (n = 5,913) of patients had their initial epilepsy diagnosis documented first in Medicare data. The initial seizure diagnosis was identified in a VA neurology clinic or by a Medicare neurology provider for approximately 27% of patients.

Clinical Characteristics

Table 3 shows comorbid conditions and the unique number of drugs received the year before epilepsy diagnosis. Consistent with a new diagnosis of epilepsy, substantial numbers had prior diagnoses of stroke, dementia, and brain

Table 2. Context of Initial Seizure Diagnosis

Variable	n (%)
Year of diagnosis	
2000	1,843 (19.0)
2001	1,978 (20.4)
2002	1,947 (20.1)
2003	2,153 (22.2)
2004	1,761 (18.2)
Source of initial diagnosis	
Neurology	2,646 (27.3)
Primary care	2,870 (29.6)
Other specialty	997 (10.3)
Hospital	1,718 (17.7)
Emergency	1,299 (13.4)
Other or missing	152 (1.6)
System of first diagnosis	
Department of Veterans Affairs	3,769 (38.9)
Medicare	5,913 (61.1)

Study cohort N = 9,682.

Table 3. Clinical Characteristics

Characteristic	n (%)
Medical comorbidities associated with epilepsy	
Cerebrovascular disease	5,538 (57.2)
Dementia	2,439 (25.2)
Brain tumor	458 (4.7)
Other neurological conditions	1,684 (17.4)
Other medical comorbidities	
Heart disease*	4,472 (46.2)
Cardiac arrhythmias	3,655 (37.8)
Hypertension	8,240 (87.0)
Diabetes mellitus	3,741 (38.5)
Renal failure	1,249 (12.9)
Liver disease	216 (2.2)
Hypercholesterolemia	5,631 (58.2)
Chronic obstructive pulmonary disease	4,337 (44.8)
Psychiatric comorbidity	
0	5,883 (60.8)
1	2,293 (23.7)
≥2	1,506 (15.6)
Unique drugs, quintile (number of drugs)	
Q-1 (0–5)	2,848 (29.4)
Q-2 (6–8)	2,533 (26.3)
Q-3 (9–11)	1,883 (19.5)
Q-4 (≥12)	2,418 (25.0)
Potential drug interaction	4,406 (45.5)

* Congestive heart failure, myocardial infarction, or angina pectoris.
Study cohort N = 9,682.

tumor. Moreover, this population had significant disease burden, evidenced by high prevalence of cardiovascular disease (myocardial infarction, angina pectoris, and congestive heart failure), cardiac arrhythmias, chronic obstructive pulmonary disease, and diseases associated with metabolic syndrome such as hypertension, diabetes mellitus, and hypercholesterolemia. The population had a high prevalence of psychiatric comorbidity diagnosed before epilepsy diagnosis, with approximately half having at least one psychiatric diagnosis recorded in administrative data. The median number of unique drugs received the year before epilepsy diagnosis was eight.

Overall, 60.1% (n = 5,822) of study subjects received a drug with potential for clinically significant interactions with the initial AED. Of those with AED-PDI exposure on the day of AED prescription, 24.0% (n = 1,416) did not receive a subsequent refill of the potentially interacting drug, leaving 4,406 (45.5%) individuals with AED-PDIs.

Table 4 shows that rates of AED-PDI were highest for phenytoin, with 57.8% of new prescriptions associated with PDI. More than half of individuals prescribed phenobarbital (51.6%) or carbamazepine (55.5%) also received a potentially interacting drug. Four groups of medications accounted for approximately 94% of those with a persistent AED-PDI: cardiovascular drugs (e.g., metoprolol, felodipine), lipid-lowering drugs (e.g., simvastatin, lovastatin), psychotropic drugs (e.g., olanzapine, trazodone), and anticoagulants (e.g., warfarin, dicoumarol). Of the cohort, 26.4% had cardiovascular drug AED-PDIs (n = 2,552), 20.7% had lipid-lowering drug AED-PDIs (n = 2,007), 14.9% had psychotropic drug AED-PDIs (n = 1,444), and 6.2% had anticoagulant

Table 4. Potential Drug Interactions(PDIs) According to Antiepileptic Drug (AED)

AED	n (%)	Initial Approach, n (%)*	Conservative Approach, n (%)†	Median Day Supply for All PDIs the Year After AED Prescription	Most Common Interacting Drugs
Phenytoin	6,477 (66.9)	4,965 (76.7)	3,721 (57.5)	239	Metoprolol, digoxin, felodipine, sertraline
Phenobarbital	246 (2.5)	167 (67.9)	127 (51.6)	265	Simvastatin, metoprolol, warfarin
Carbamazepine	824 (8.5)	565 (68.6)	457 (55.5)	216	Simvastatin, metoprolol, warfarin
Valproate	545 (5.6)	119 (21.8)	97 (17.0)	232	Sertraline, paroxetine, amitriptyline,
Totals	9,682 (100)	5,822 (60.1)	4,406 (45.5)		

AEDs with fewer than 11 potential drug interactions are not presented because of Department of Veterans Affairs data security requirements.
* All individuals who received an AED with a clinically meaningful drug interaction with ongoing medication on the day of the initial AED prescription.
† Only individuals with a subsequent prescription of interacting medications.

AED-PDIs (n = 601). Of those with a persistent AED-PDI, 48.0% had two or more concomitant AED-PDIs.

The duration of exposure was often long. Table 4 shows the median number of days of PDI prescription for the first year of epilepsy treatment. The day supply for AED-PDI was more than 200 days, with medians ranging from 216 to 265 days.

Odds ratios (ORs) and 99% confidence intervals (CIs) from the multivariable logistic regression analysis predicting AED-PDI exposure are presented in Table 5. The strongest predictors of PDI were having hypertension or hypercholesterolemia and receiving six or more medications the year before epilepsy diagnosis. Moreover, individuals initially diagnosed with epilepsy in an emergency department or primary care settings were significantly more likely to have AED-PDI exposure than were those initially diagnosed in a neurology setting. Individuals with a lower likelihood for AED-PDI exposure included those who were older and those with a diagnosis of liver disease.

Logistic regression analysis predicting renewal of interacting drug after initiation of AED (full data not shown) found that age, several chronic disease states, and year of initial AED prescription were significant predictors. Those who were aged 66 to 74 were less likely to have an interacting drug renewed after the prescription of the initial AED than those who were aged 75 to 84 (OR = 0.82, 99% CI = 0.70–0.98, $P = .004$). Individuals with dementia (OR = 0.80, 99% CI = 0.66–0.97, $P = .002$), metastatic cancer (OR = 0.64, 99% CI = 0.47–0.86, $P < .001$), and chronic obstructive lung disease (OR = 0.72, 99% CI = 0.60–0.85, $P < .001$) were less likely to have interacting drugs renewed. Finally, those receiving their AED in FY04 were less likely to have the interacting drug renewed after the prescription of the initial AED than were individuals who received their AED in FY00 (OR = 0.68, 99% CI = 0.52–0.88, $P < .001$).

Table 5. Logistic Regression: Predictors of Potential Antiepileptic Drug–Drug Interaction

Variable	Odds Ratio (99% Confidence Interval)	*P*-Value
Demographic characteristic		
Age (vs 66–74)		
75–84	0.86 (0.77–0.96)	<.001
≥85	0.69 (0.53–0.91)	<.001
Race (vs white)		
Black	1.02 (0.87–1.19)	.72
Hispanic	1.00 (0.77–1.29)	.98
Other	0.87 (0.48–1.58)	.55
Female (vs male)	0.79 (0.53–1.19)	.14
Married (vs not married)	1.11 (0.98–1.24)	.03
Context of epilepsy diagnosis		
Year (vs 2000)		
2001	0.91 (0.77–1.08)	.16
2002	0.89 (0.75–1.06)	.08
2003	0.85 (0.72–1.01)	.02
2004	0.83 (0.69–0.99)	.01
Diagnosis in Medicare (vs diagnosis in Department of Veterans Affairs)	1.02 (1.01–1.04)	<.01
Setting of epilepsy diagnosis (vs neurology)		
Emergency department	1.30 (1.08–1.58)	<.001
Hospital	1.08 (0.91–1.28)	.26
Other specialty care	1.20 (1.00–1.45)	.01
Primary care	1.29 (1.12–1.49)	<.001
Clinical characteristics		
Cerebrovascular disease	1.22 (1.08–1.36)	<.001
Dementia	0.94 (0.82–1.07)	.22
Brain tumor	0.86 (0.66–1.10)	.08
Other neurological conditions	1.01 (0.87–1.17)	.86
Cardiovascular disease*	1.04 (0.92–1.17)	.86
Cardiac arrhythmia	1.05 (0.93–1.18)	.31
Hypertension	1.46 (1.24–1.82)	<.001
Diabetes mellitus	0.91 (0.81–1.03)	.04
Renal failure	0.91 (0.77–1.07)	.13
Liver disease	0.70 (0.50–0.98)	.01
Hypercholesterolemia	1.40 (1.24–1.57)	<.001
Chronic obstructive pulmonary disease	0.91 (0.81–1.11)	.03
Psychiatric conditions: one or more (vs none)	0.99 (0.88–1.11)	.76
High medication burden (≥6 medications)	1.36 (1.20–1.55)	<.001

N for multivariate analysis = 9,624.
* Congestive heart failure, myocardial infarction, or angina pectoris.

DISCUSSION

Despite the availability of AEDs with low potential for AED-PDI, the majority of older patients with epilepsy in the United States are treated with AEDs with high AED-PDI potential (phenytoin, carbamazepine, and phenobarbital).[11,13] The problem of AED-PDI has been identified as a consideration in selecting AEDs for older patients in numerous review articles.[3–6] To the authors' knowledge, this is the first study that has systematically examined the scope of the AED-PDI problem in a geriatric population and further identified the patient characteristics (diagnostic setting, age, comorbidities) associated with risk of AED-PDI. Nearly half of all older patients in the current study received an AED that interacted with existing, chronic medications. Of those with AED-PDI exposure, nearly half had multiple AED-PDIs, and AED-PDI exposure was long, suggesting that healthcare providers need to increase their awareness of AED-PDIs in caring for older adults.

Of the four classes of concomitant medications that accounted for 94% of AED-PDIs, the most common were drugs used in the treatment of vascular disease. This is not surprising, because stroke is the most common etiology associated with new-onset epilepsy in older adults, and 80% of older adults have been diagnosed with hypertension and hyperlipidemia.[2,13] These potential interactions are of great clinical relevance, because they may affect health outcomes. Potential interactions identified by this study may significantly reduce the efficacy of antihypertensive medications, including beta-blockers, calcium agonists, and angiotensin-receptor blockers, in some case without altering serum drug concentrations.[25–28] In addition, previous research has found that co-administration of carbamazepine with orally

administered simvastatin, a 3-hydroxy-3-methylglutaryl coenzyme A reductase inhibitor used in the treatment of dyslipidemia, led to an 80% reduction in bioavailability of simvastatin.[29] This effect would require increasing the dose of simvastatin by a factor of four to achieve therapeutic blood concentrations. Lastly, warfarin-related AED-PDIs may result in altered anticoagulation properties of this drug, leading to a prolongation of the prothrombin time. All of these AED-PDIs could potentially increase the risk for stroke and heart attack and hemorrhage, leading to significant patient morbidity and mortality.

These data highlight the complexity of adding medications to the regimens of older patients with multiple medical conditions. Healthcare providers caring for older patients must be vigilant in the selection of an AED, especially in those with vascular disease, and closely monitor the treatment of hypertension, dyslipidemia, and anticoagulation therapy. Although the data did not allow this clinical behavior to be examined, future research should consider the extent to which such monitoring is conducted.

The examination of risk factors for AED-PDIs and the likelihood of having an interacting drug renewed suggest that clinicians in neurology clinics may have a greater awareness of AED-PDIs than providers in primary care or the emergency department. The data also suggested more caution on the practioner's part with regard to medication problems in patients who might be considered frail. The oldest-old (85) and individuals with dementia and liver disease were less likely to have AED-PDI exposure than their younger counterparts and those without such disease. The finding regarding age is consistent with previous findings from studies of suboptimal prescribing in older adults.[30–32] The consistency of these findings to more general studies of suboptimal prescribing in older adults suggest that providers are generally more cautious when making changes to medication regimens in these more-vulnerable patients. Regarding liver disease, it is possible that providers preferentially prescribed gabapentin—a drug that is primarily renally cleared and has no significant drug interactions—for those with impaired liver function.[33]

The findings of the current study must be interpreted in light of several unique features of the study. First, it examined only AED-PDIs in those with new-onset epilepsy. This "incident case" approach allowed the extent to which providers are aware of, and consider, AED-PDIs in their initial AED prescribing decision and early on in treatment to be better understood, because AEDs are rarely changed once seizures are successfully controlled unless the patient experiences adverse drug events. Although individuals who did not have interacting medications continued after the initial AED prescription were excluded from the analysis, the method did not examine changes made to prescriptions that were continued. It is possible that clinicians altered dosage or increased monitoring in response to the AED-PDI. Future research should examine this possibility.

A second feature is that this study was based in a single integrated healthcare system (the VA healthcare system) in which most of the cohort had a copayment waiver because of poverty or severe disability. Thus, the generalizability of the findings to other settings is unclear. Other systems with a fee-for-service model or multitiered formulary may have had higher rates of AED-PDIs during the study period because AEDs with fewer PDIs tended to be more expensive than phenytoin or carbamazepine. Because prescribing patterns for AEDs appear to be similar to those presented previously[11] and this cohort included those whose first diagnosis (and possibly AED prescription) occurred in the context of a non-VA Medicare provider, it is likely that this bias is minimized, although it is possible that AED-PDI patterns are different in women because studies have found different patterns of suboptimal prescribing in women,[34] so further exploration of AED-PDIs in older women is needed.

Moreover, the analysis included only medications dispensed by the VA, but during the time of this study, most older veterans who receive prescription drugs through VA received few prescription drugs from outside sources;[35] thus this potential bias is limited.

Finally, these data reflect prescribing patterns several years ago. Increasing use of automatic, computerized medication alerts at the time of prescribing within the VA have occurred since this time and may be reflected in the significant reduction of AED-PDIs between October 1, 1999, and September 30, 2004, and the finding that individuals with AED-PDIs were more likely to have the interacting drug dropped in FY04 than in FY00. Examination of more-contemporaneous data is needed to examine changes in AED patterns and concomitant changes in AED-PDIs.

These data suggest that exposure to AED-PDI is highly prevalent in older patients newly diagnosed with epilepsy, primarily because of the high rates of use of phenytoin and to a lesser extent carbamazepine. Despite evidence of slowly growing awareness of this patient safety problem, there is substantial room for improvement. Because pharmacological data suggest that AED-PDIs may reduce the efficacy of the interacting drug or cause AED toxicity, prolonged AED-PDI exposure has the potential to result in adverse outcomes such as falls and fractures, stroke, myocardial infarction, excessive anticoagulation, or even death.[6,7] Thus, it is time to consider implementing interventions to reduce AED-PDIs in older patients with epilepsy. Data from prior studies suggest that integrating neurology[36] and clinical pharmacy consultation,[37] combined with electronic medical records systems that alert clinicians to PDIs,[38] may improve care quality and outcomes in this vulnerable population.

ACKNOWLEDGMENTS

This study was funded by the Department of Veterans Affairs, Health Services Research and Development Service (IIR-02-274, Dr. Pugh PI) and the Epilepsy Foundation (Dr. Pugh PI). Dr. Copeland received funding from VA Health Services Research and Development Service (HSR&D) Merit Review Entry Program Award (MRP-05-145). Dr. Steinman was supported by a VA HSR&D Career Development Transition Award (01-013) and a National Institute on Aging Paul Beeson Career Development Award (1K23AG030999). The authors also acknowledge and appreciate support from the South Texas Veterans Healthcare System/Audie L. Murphy Division and the VERDICT research program. The views expressed in this article are those of the authors and do not necessarily represent the views of the VA. We also acknowledge contributions of other members of the TIGER research team: Joyce Cramer

BA, Jeffrey Tabares BA, Francesca Cunningham PharmD, Omotola Hope MD, Nancy Kressin PhD, Barbara Bokhour PhD, Mark Glickman, PhD, Richard Lewis, MSA.

Conflict of Interest: Anne C. Van Cott has research funding from GlaxoSmithKline. Michael A. Steinman served as an expert witness for the plaintiff in *United States ex. rel. Franklin v Pfizer, Inc.*, which alleged that Parke-Davis, Inc. illegally marketed Neurontin (gabapentin) for uses not approved by the Food and Drug Administration. The remaining authors report no conflicts of interest.

Author Contributions: Mary Jo Pugh: study concept and design, funding, analysis, interpretation and preparation of manuscript. Anne C. Van Cott, Eric M. Mortensen, Janice E. Knoefel, and Dan R. Berlowitz: interpretation of data, and preparation of manuscript. Michael A. Steinman and Chen-Pin Wang: study design, interpretation of data, preparation of manuscript. Megan E. Amuan Acquisition of data, analysis, and preparation of manuscript.

Sponsor's Role: The study sponsors, VA Health Services Research and Development and the Epilepsy Foundation, had no role in this study's design or conduct.

REFERENCES

1. Hauser WA. Seizure disorders: The changes with age. Epilepsia 1992;33: S6–S14.
2. Rowan AJ, Ramsay RE, Collins JF et al. New onset geriatric epilepsy: A randomized study of gabapentin, lamotrigine, and carbamazepine. Neurology 2005;64:1868–1873.
3. Leppik IE, Birnbaum A. Epilepsy in the elderly. Semin Neurol 2002;22: 309–320.
4. Willmore LJ. Antiepileptic drug therapy in the elderly. Pharmacol Ther 1998;78:9–16.
5. Rowan AJ. Seizure. Fundamentals of drug management of epilepsy in the older patient. Geriatrics 2002;57:33–37; quiz 38.
6. Patsalos PN, Perucca E. Clinically important drug interactions in epilepsy: Interactions between antiepileptic drugs and other drugs. Lancet Neurol 2003;2:473–481.
7. Patsalos PN, Froscher W, Pisani F et al. The importance of drug interactions in epilepsy therapy. Epilepsia 2002;43:365–385.
8. Mancia G. Prevention and treatment of stroke in patients with hypertension. Clin Ther 2004;26:631–648.
9. Bejot Y, Ben Salem D, Osseby GV et al. Epidemiology of ischemic stroke from atrial fibrillation in Dijon, France, from 1985 to 2006. Neurology 2009;72:346–353.
10. Delahoy PJ, Magliano DJ, Webb K et al. The relationship between reduction in low-density lipoprotein cholesterol by statins and reduction in risk of cardiovascular outcomes: An updated meta-analysis. Clin Ther 2009;31:236–244.
11. Gidal BE, French JA, Grossman P et al. Assessment of potential drug interactions in patients with epilepsy: Impact of age and sex. Neurology 2009;72:419–425.
12. Holden EW, Grossman E, Nguyen HT et al. Developing a computer algorithm to identify epilepsy cases in managed care organizations. Dis Manage 2005;8:1–14.
13. Pugh MJ, Van Cott AC, Cramer JA et al. Trends in antiepileptic drug prescribing for older patients with new-onset epilepsy: 2000–2004. Neurology 2008;70:2171–2178.
14. Micromedic. Drug-Reax* system. In: Gelman CR, Rumack BH, editors. Micromedex Intranet Knowledge Bases. Denver, CO: Micromedex, Inc., 2003.
15. Elixhauser A, Steiner C, Harris DR et al. Comorbidity measures for use with administrative data. Med Care 1998;36:8–27.
16. 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.
17. Krishnan LL, Petersen NJ, Snow AL et al. Prevalence of dementia among Veterans Affairs medical care system users. Dement Geriatr Cogn Disord 2005;20:245–253.
18. 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.
19. So EL, Annegers JF, Hauser WA et al. Population-based study of seizure disorders after cerebral infarction. Neurology 1996;46:350–355.
20. Dhanuka AK, Misra UK, Kalita J. Seizures after stroke: A prospective clinical study. Neurol India 2001;49:33–36.
21. Amatniek JC, Hauser WA, Delcastillo-Castaneda C et al. Incidence and predictors of seizures in patients with Alzheimer's disease. Epilepsia 2006;47:867–872.
22. Lozsadi DA, Larner AJ. Prevalence and causes of seizures at the time of diagnosis of probable Alzheimer's disease. Dement Geriatr Cogn Disord 2006;22:121–124.
23. 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.
24. Centers for Disease Control and Prevention (CDC). Centers for Disease Control and Prevention (CDC). The State of Aging and Health in America 2007. Whitehouse Station, NJ: The Merck Company Foundation, 2007.
25. Flockhart DA, Tanus-Santos JE. Implications of cytochrome P450 interactions when prescribing medication for hypertension. Arch Intern Med 2002;162:405–412.
26. Michelucci R, Cipolla G, Passarelli D et al. Reduced plasma nisoldipine concentrations in phenytoin-treated patients with epilepsy. Epilepsia 1996;37:1107–1110.
27. Fischer TL, Pieper JA, Graff DW et al. Evaluation of potential losartan-phenytoin drug interactions in healthy volunteers. Clin Pharmacol Ther 2002;72:238–246.
28. Bahls FH, Ozuna J, Ritchie DE. Interactions between calcium channel blockers and the anticonvulsants carbamazepine and phenytoin. Neurology 1991;41:740–742.
29. Ucar M, Neuvonen M, Luurila H et al. Carbamazepine markedly reduces serum concentrations of simvastatin and simvastatin acid. Eur J Clin Pharmacol 2004;59:879–882.
30. Piecoro LT, Browning SR, Prince TS et al. A database analysis of potentially inappropriate drug use in an elderly medicaid population. Pharmacotherapy 2000;20:221–228.
31. Goulding MR. Inappropriate medication prescribing for elderly ambulatory care patients. Arch Intern Med 2004;164:305–312.
32. 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.
33. Willmore LJ. Choice and use of newer anticonvulsant drugs in older patients. Drugs Aging 2000;17:441–452.
34. Bierman AS, Pugh MJ, Dhalla I et al. Sex differences in inappropriate prescribing among elderly veterans. Am J Geriatr Pharmacother 2007;5:147–161.
35. Kaboli PJ, McClimon BJ, Hoth AB et al. Assessing the accuracy of computerized medication histories. Am J Manage Care 2004;10:872–877.
36. Hope OA, Zeber JE, Kressin NR et al. New-onset geriatric epilepsy care: Race, setting of diagnosis, and choice of antiepileptic drug. Epilepsia 2009;50:1085–1093.
37. Cowper PA, Weinberger M, Hanlon JT et al. The cost-effectiveness of a clinical pharmacist intervention among elderly outpatients. Pharmacotherapy 1998;18:327–332.
38. Terrell KM, Perkins AJ, Dexter PR et al. Computerized decision support to reduce potentially inappropriate prescribing to older emergency department patients: A randomized, controlled trial. J Am Geriatr Soc 2009;57:1388–1394.

A Pilot Study Assessing Knowledge of Clinical Signs and Physical Examination Skills in Incoming Medicine Residents

Subha Ramani, MBBS, MMEd, MPH
Brandi N. Ring, MA
Robert Lowe, MD
David Hunter, MBBS, MSc, PhD

Abstract

Background Physical exam skills of medical trainees are declining, but most residencies do not offer systematic clinical skills teaching or assessment.

Objective To assess knowledge of clinical signs and physical exam performance among incoming internal medicine residents.

Method For this study, 45 incoming residents completed a multiple choice question test to assess knowledge of clinical signs. A random selection of 20 underwent a faculty-observed objective structured clinical examination (OSCE) using patients with abnormal physical findings. Mean percentage scores were computed for the multiple choice question test, overall OSCE, and the 5 individual OSCE systems.

Results The mean scores were 58.4% (14.6 of 25; SD 11. 5) for the multiple choice question test and 54.7% (31.7 of 58; SD 11.0) for the overall OSCE. Mean OSCE scores by system were cardiovascular 30.0%, pulmonary 69.2%, abdominal 61.6%, neurologic 67.0%, and musculoskeletal 41.7%. Analysis of variance showed a difference in OSCE system scores ($P < .001$) with cardiovascular and musculoskeletal scores significantly lower than other systems.

Conclusion Overall, physical exam knowledge and performance of new residents were unsatisfactory. There appears to be a pressing need for additional clinical skills training during medical school and residency training and we are planning a new clinical skills curriculum to address this deficiency.

Editor's Note: The online version of this article includes the sample questions for the multiple choice test and the faculty scoring sheet for the OSCE for the cardiac exam station.

Subha Ramani, MBBS, MMEd, MPH, is Associate Program Director, Internal Medicine Residency Program, Associate Professor of Medicine, Boston University School of Medicine; **Brandi N. Ring, MA,** is Medical Student, Boston University School of Medicine; **Robert Lowe, MD,** is Education Director, Division of Gastrenterology, Department of Medicine, Evans Educator for Professionalism, and Associate Professor of Medicine, Boston University School of Medicine; and **David Hunter, MBBS, MSc, PhD,** is Professor and ARC Future Fellow, School of Medicine, Northern Clinical School, University of Sydney, Australia.

We gratefully acknowledge the support received from the General Clinical Research Center Grant M01 RR00533 for allowing us to use their facility for the OSCE testing.

We are indebted to the residents for participating in the study, the Internal Medicine Residency Program Office staff for helping to organize the OSCE, Department of Medicine Faculty who served as the teachers and examiners, and, most importantly, our patients who enthusiastically participated in the study for the sole purpose of educating our trainees.

Corresponding author: Subha Ramani, MBBS, MMEd, MPH, Boston University School of Medicine, 72, East Concord Street, Evans 124, Boston, MA 02118, 617.638.7985, Subha.ramani@bmc.org

Received December 26, 2009; revision received February 19, 2010; accepted April 22, 2010.

DOI: 10.4300/JGME-D-09-00107.1

Background

Physical examination skills traditionally have been viewed among the most valuable skills taught during medical education,[1–4] contributing to more cost-effective use of diagnostic services, while rewarding physicians with the excitement and satisfaction of making a diagnosis using their knowledge and skills.[1,2] These skills also increase direct contact with patients, and the therapeutic value of the human touch is impossible to quantify.[2]

Several investigators have reported an overall decline in clinical skills of medical students and residents,[5–10] with residents less well prepared for taking an adequate medical history, performing a reliable physical examination, and effectively communicating with patients,[11] while relying on ordering tests without always knowing how to interpret them.[1,11] In an era of increasing health care costs we need to reconsider the importance of physical examination skills.[1,2]

Despite documented deficiencies in clinical skills, medical school and residency curricula do not emphasize clinical skills teaching or assessment.[3,7,12] Reported barriers to teaching clinical exam skills include a scarcity of good teaching patients, lack of time for teaching at the bedside, an over-reliance on technology, and a shortage of skilled faculty to impart this knowledge.[11,13]

Before developing a new clinical skills curriculum for our internal medicine residency program, we wished to explore the physical exam skills of our incoming residents as needs assessment.

Our study objectives were

1. To investigate the knowledge of clinical signs as well as the physical exam skills of new postgraduate year 1 (PGY-1) residents using volunteer patients during an objective structured clinical examination (OSCE)
2. To explore system-specific strengths and weaknesses in their physical exam skills

Methods

Setting and Participants

Incoming internal medicine PGY-1 residents at Boston University School of Medicine completed a written multiple choice question (MCQ) test on clinical signs and a randomly selected subsample completed a physical exam assessment during their residency orientation in June 2006. These tests were designed as a pretest prior to implementation of a new clinical skills curriculum. The protocol for this study was approved by the Institutional Review Board at Boston University School of Medicine.

A planning committee consisting of generalist and subspecialist faculty from the Department of Medicine discussed and finalized the MCQ test questions and the OSCE scoring sheets after review of questions from the Membership of the Royal College of Physicians and other examinations as well as detailed discussions of essential elements of system-specific physical examination. This core group of faculty consistently taught residents bedside clinical skills, had a reputation of being skilled clinical diagnosticians, and served as preceptors for the OSCE.

The written test consisted of 25 MCQs designed to evaluate the ability of the residents to interpret and diagnose physical exam findings. The 5 major systems, cardiovascular, neurologic, pulmonary, gastrointestinal, and musculoskeletal, were represented.

The OSCE used volunteer patients with abnormal physical findings recruited from the medical wards and clinics. The 5 stations included cardiac, pulmonary, abdominal, neurologic, and musculoskeletal systems. At each station a faculty examiner instructed the residents to perform a focused physical exam, completed a scoring sheet, and provided feedback in 10 minutes. All faculty preceptors underwent an orientation to the procedures of observation and feedback during the OSCE as well as the scoring. A sample score sheet is included in Appendix B. Each item on the scoring sheet was graded on a 3-point rating scale from 0 to 2 points. Two points were awarded if the element was performed correctly, 1 point if the element was performed with room for improvement, and no points awarded if the exam was omitted. The OSCE encompassed 5 to 6 elements on exam technique and 1 to 2 elements on interpretation and diagnosis, for a maximum score of 58 points.

Statistical Analysis

Descriptive analysis was performed on the data collected for the PGY-1 cohort, and we calculated mean percentage scores for the MCQ test and the OSCE. The OSCE scores were then analyzed by individual organ system. We wished to examine whether there were differences in physical exam performance for different systems. Scores for each individual system were ranked and then compared using analysis of variance. If a difference was detected, we explored further where the differences lay by pair-wise comparison using Tukey minimum significant differences. All analysis was run at $\alpha = .05$ level using SAS version 9.0 (SAS Inc, Cary, NC) and Excel XP (Microsoft, Redmond, WA).

Results

A total of 45 internal medicine PGY-1 residents at Boston University School of Medicine completed the MCQ test and 20 completed the 5-station physical exam OSCE. Most were US medical graduates from several different medical schools; 2 were international medical graduates.

The overall mean score for the written test was 58.4% with a standard deviation of 11.5 and range 36.0 to 80.0 ($n = 45$). There was no statistical difference between the scores of those residents chosen to undergo the OSCE and those not selected ($P = .261$).

The mean overall OSCE score was 54.7% with a standard deviation of 11.0 and range 39.7 to 84.5 ($n = 20$). Analysis of variance showed significant differences in the OSCE scores for individual systems ($P < .05$) with the cardiovascular and musculoskeletal examination scores being significantly lower than the pulmonary, neurology, and abdominal examinations scores. The overall MCQ, OSCE, and individual system scores are shown in the TABLE.

Examples of errors observed in residents' physical exam include

1. Faulty exam technique
   a. Not using bell and diaphragm of stethoscope
   b. Not eliciting shifting dullness correctly
2. Lack of systematic exam
   a. Skipping inspection or palpation completely
   b. Not following a stepwise exam such as motor strength, tone, reflexes, gait, and so forth
3. Failure to identify findings
   a. Identification of diastolic murmur
   b. Identification of bronchial breath sounds
4. Failure to interpret findings and make a diagnosis
   a. Differentiating between upper and lower motor neuron signs

TABLE **Mean Percent Scores**

Test	N	Mean	Median	SD	Score Range
MCQ (all PGY-1)	45	58.4	56.0	11.5	36.0–80.0
MCQ (PGY-1 w/OSCE)	20	61.0	60.0	11.9	36.0–80.0
OSCE total	20	54.7	61.3	11.0	39.7–84.5
Cardiology	20	30.0	30.0	21.3	0–100
Pulmonary	20	69.2	66.6	18.2	41.7–100
Abdominal	20	61.6	58.3	16.6	33.3–91.7
Neurologic	20	67.0	66.6	14.8	33.3–100
Musculoskeletal	20	41.7	37.5	17.1	25–100

Abbreviations: MCQ, multiple choice question; OSCE, objective structured clinical examination; PGY-1, postgraduate year 1.

5. Difficulty formulating differential diagnosis for a given finding
   a. Causes of ascites
   b. Causes of knee effusions

Discussion

The newly graduated medical students entering internal medicine residency in our study scored less than 60% on average in a knowledge test of clinical signs as well as a physical exam OSCE. Errors were noted in physical exam technique as well as diagnosis. Many studies have reported less than satisfactory physical exam skills among trainees. Dupras and Li[14] found mean scores of 50 ± 11 in physical examination and diagnosis stations of an OSCE assessing postgraduate year 2 internal medicine residents.[14] Vukanovic-Criley and colleagues[9] reported that cardiac exam skills improved between years 1 and 2 of medical school and reached a plateau thereafter.

The variability of exam performance between systems may be a result of different exposures during medical school clinical rotations and a measure of interest in the given field. However, based on the substantial difference from other systems we surmise that there are deficiencies in both cardiovascular and musculoskeletal teaching.

The strengths of this study are the use of trained faculty members to observe and assess the OSCE allowing for more consistent scoring. The study used real patients with abnormal physical examination findings rather than standardized patients with scripted history and answers to questions. This is a more accurate representation of the kind of encounters that medical trainees will be exposed to in clinical practice. The OSCE also served as a learning tool as participants were provided with immediate feedback from faculty on their physical examinations skills.

Limitations

Our study has some limitations. We tried to limit variability in scoring by faculty orientation and use of objective and easily observable physical exam behaviors, but different faculty observers may exhibit variable leniency in scoring. There also may have been variability in the degree of difficulty between patients at different stations. We present cross-sectional data that may reflect many different factors including prior medical school training and interest. We hope to ascertain in subsequent studies the trajectory of knowledge and skills following implementation of a clinical skills curriculum. We did not set out to test the psychometric properties of our assessment instrument but tried to address content validity to ensure that the instrument is appropriate to measure what we intended to measure, but future studies need to test the validity of instruments that can assess physical diagnosis skills with precision while being reliable.

Conclusions

There is a strong indication that additional clinical skills training and assessment are needed during medical school as well as residency training. Standardized patient examinations should supplement, but not replace, direct observation of trainees as only real patient interactions can help educators document longitudinal growth of trainees' clinical skills.[14] Faculty development is also crucial as many clinical faculty do not feel confident of their own clinical skills.[13] Questions have been raised about the validity of standardized patient encounters alone in assessing higher-level trainees as they cannot assess diagnostic skills.[15–17]

Although many educators espouse the value of clinical skills,[1–3,18] some skeptics have questioned whether these skills are anachronistic.[19] Future studies will need to investigate whether a formal curriculum will improve and sustain the clinical skills of residents. The predictive values of many physical examination techniques and findings have

been evaluated,[2,18] and the *JAMA* Rational Clinical Examination series is the foremost example of evidence-based physical diagnosis.[20] It is vital that clinical educators continue to study the clinical utility of physical findings and discard signs or maneuvers of little value; modern technology can be very helpful in achieving these goals.

Our pilot study examined the status of physical exam skills in a sample of newly graduated medical students. It is one step in a comprehensive attempt to assess the details of physical examination deficiencies among our residents to advance our understanding of the underlying concepts of such deficiencies. Ideally, before the findings are used in the design of a clinical skills curriculum, a larger group of residents needs to be studied using precurriculum and postcurriculum intervention assessment, and the reliability and validity of instruments to assess physical diagnosis skills need to be tested, to assess whether previously validated OSCE assessments would be useful. To reinforce and maintain the improvement in clinical skills that could be achieved through a systematic curriculum, educators also need to reinforce these skills in the clinical setting, role-model their use, and demonstrate their value in quality patient care. Finally, research is needed to assess whether improvement in clinical skills is essential for patient care in modern-day medicine, namely, whether improved clinical skills lead to more timely diagnosis, reductions in inappropriate use of resources, and improvements in patient satisfaction.

References

1 Bordage G. Where are the history and the physical? *CMAJ*. 1995;152:1595–1598.
2 Mangione S, Peitzman SJ. Physical diagnosis in the 1990s: art of artifact? *J Gen Intern Med*. 1996;11:490–493.
3 Kern DC, Parrino TA, Korst DR. The lasting value of clinical skills. *JAMA*. 1985;254:70–76.
4 Holmboe ES. Faculty and the observation of trainees' clinical skills: problems and opportunities. *Acad Med*. 2004;79:16–22.
5 Wiener S, Nathanson M. Physical examination: frequently observed errors. *JAMA*. 1976;236:852–855.
6 Mangione S, Burdick WP, Peitzman SJ. Physical diagnosis skills of physicians in training: a focused assessment. *Acad Emerg Med*. 1995;2:622–629.
7 Mangione S, Nieman LZ. Cardiac auscultatory skills of internal medicine and family practice trainees: a comparison of diagnostic proficiency. *JAMA*. 1997;278:717–722.
8 Mangione S, Nieman LZ. Pulmonary auscultatory skills during training in internal medicine and family practice. *Am J Respir Crit Care Med*. 1999;159:1119–1124.
9 Vukanovic-Criley JM, Criley S, Warde CM, et al. Competency in cardiac examination skills in medical students, trainees, physicians, and faculty: a multicenter study. *Arch Intern Med*. 2006;166:610–616.
10 Johnson JE, Carpenter JL. Medical house staff performance in physical examination. *Arch Intern Med*. 1986;146:937–941.
11 Fred HL. Hyposkillia: deficiency of clinical skills. *Tex Heart Inst J*. 2005;32:255–257.
12 Mangione S, Duffy FD. The teaching of chest auscultation during primary care training: has anything changed in the 1990s? *Chest*. 2003;124:1430–1436.
13 Ramani S, Orlander JD, Strunin L, Barber TW. Whither bedside teaching?: a focus-group study of clinical teachers. *Acad Med*. 2003;78:384–390.
14 Dupras DM, Li JT. Use of an objective structured clinical examination to determine clinical competence. *Acad Med*. 1995;70:1029–1034.
15 Kopelow ML, Schnabl GK, Hassard TH, et al. Assessing practicing physicians in two settings using standardized patients. *Acad Med*. 1992;67:S19–S21.
16 Ram P, van der Vleuten C, Rethans JJ, Grol R, Aretz K. Assessment of practicing family physicians: comparison of observation in a multiple-station examination using standardized patients with observation of consultations in daily practice. *Acad Med*. 1999;74:62–69.
17 Rethans JJ, Sturmans F, Drop R, van der Vleuten C, Hobus P. Does competence of general practitioners predict their performance?: comparison between examination setting and actual practice. *BMJ*. 1991;303:1377–1380.
18 Feddock CA. The lost art of clinical skills. *Am J Med*. 2007;120:374–378.
19 Jauhar S. The demise of the physical exam. *N Engl J Med*. 2006;354:548–551.
20 Sackett DL The rational clinical examination: a primer on the precision and accuracy of the clinical examination. *JAMA*. 1992;267:2638–2644.

Cohort study

High blood pressure while taking antithrombotic medication is associated with an increased risk of developing intracranial haemorrhage

Adam J Rose, Elaine M Hylek

10.1136/ebm1123

Centre for Health Quality, Outcomes, and Economic Research, Bedford VA Medical Center and Boston University School of Medicine, Bedford, Massachusetts, USA

Correspondence to:
Adam J Rose
Center for Health Quality, Outcomes, and Economic Research, Bedford VA Medical Center, 200 Springs Road, Building 70, Bedford, MA 01730, USA;
adamrose@bu.edu

Commentary on: **Toyoda K**, Yasaka M, Uchiyama S, *et al.*; Bleeding with Antithrombotic Therapy (BAT) Study Group. Blood pressure levels and bleeding events during antithrombotic therapy: the Bleeding with Antithrombotic Therapy (BAT) Study. *Stroke* 2010;**41**:1440–4.

Context

Major haemorrhage, particularly intracranial haemorrhage (ICH), is a dreaded consequence of antithrombotic therapy. Strategies to mitigate this risk are needed as indications for antithrombotic therapy increasingly extend to older, high-risk patients. A previous large-scale randomised controlled trial demonstrated the benefits of lowering blood pressure (BP) in preventing recurrent stroke.[1] In addition, a posthoc analysis of pooled data from the SPORTIF trials suggested a threshold BP of <140/80 mm Hg to decrease risk of stroke. In that study, extrapolation to haemorrhagic stroke was limited by the small number of ICH events (n=17).[2] The current study sought to validate these findings in a non-trial setting and to better define the temporal relationship between BP and recurrent stroke.

Methods

The study focused on a prospective, observational cohort of 4009 patients from 19 stroke and cardiovascular centres in Japan. All patients were receiving antiplatelet agents and/or warfarin for cardiovascular prevention; most (57%) had a prior ischemic or haemorrhagic stroke. Patients were observed for 2–30 months; BP was measured at 1-month intervals. The outcome of interest, assessed by questionnaire at each monthly visit, was major haemorrhage. For this study, patients were divided into three groups: ICH, ECH (extracranial haemorrhage requiring transfusion or entailing a risk of disability) and no haemorrhage. The relationship between BP (both systolic (SBP) and diastolic (DBP)) and outcomes was analysed using entry BP, mean overall BP and most recent BP. Covariates (known or suspected risk factors for haemorrhage) included age, sex, prior stroke, heart disease, cancer, cirrhosis, hypertension, diabetes, hyperlipidemia, current or previous smoking and alcohol consumption of two or more drinks per day.

Findings

BP at study entry was not a significant predictor of outcomes after adjustment. However, after adjustment, average SBP between 1 and 6 months of follow-up (HR 1.45 per 10 mm/Hg increase), average SBP between 7 and 12 months (HR 1.47) and average DBP between 7 and 12 months (HR 2.05) were significantly associated with ICH (but not ECH). The relationship between the most recent BP and the risk of ICH was even stronger and more monotonic: BP preceding ICH (142/81 mm/Hg) was significantly higher than BP preceding no haemorrhage (132/75 mm/Hg). Using receiver operating characteristic curves, the optimal BP level to reduce the risk of ICH was determined to be 130/81 mm/Hg or lower.

Commentary

This study affirms the relationship between BP and risk of ICH in a non-trial, high-risk population receiving antithrombotic therapy. As older age, prior stroke and antithrombotic therapy all increase the risk of ICH, the importance of BP control in this patient population cannot be overemphasised. It is of note that in this study, BP worsened over time, unlike the improved control seen in most trials. This would have strengthened the association between the most recent BP and ICH. More intensive management of BP can improve control,[3] and it is likely that the patients in this study would have benefited from more intensive BP management. As acknowledged by the investigators, the relatively small number of ICH events (n=31) precluded exploration of BP thresholds among the highest-risk subsets, for example, older individuals on dual therapy. In addition, almost half of the patients were taking warfarin; it would have been ideal to account for the contribution of INR (international normalized ratio) control as a predictor of major haemorrhage[4] when examining the impact of BP.

This study also showed that while uncontrolled BP increases the risk of ICH, it does not seem to increase the risk of ECH. It is likely that extracranial vascular beds are less subject to the changes in flow and sheer stress and subsequent vascular remodelling wrought by uncontrolled hypertension. The authors properly suggest that though scrupulous BP control may help reduce the risk of ICH, reducing the risk of ECH may require other approaches.

Overall, this study demonstrates that real clinical benefit can be expected from efforts to improve BP

control in high-risk patients. The next step would be an implementation study of a programme to improve BP control in patients receiving antiplatelet and/or anticoagulant therapy. Such a study should be designed to demonstrate not only that the intervention helped patients but also how it can practicably be replicated in diverse settings.

Acknowledgements AJR is supported by a Career Development Award from the Veterans Affairs Health Services Research and Development Service. The opinions expressed in this manuscript do not necessarily represent the official views of the Department of Veterans Affairs.

Competing interests AJR has no competing interests to report. EMH has received honoraria from Bayer and Bristol Myers Squibb and is on the advisory boards for Boehringer-Ingelheim, Bristol Myers Squibb, Merck and Sanofi Aventis.

References

1. **Chapman** N, Huxley R, Anderson C, *et al*. Effects of a perindopril-based blood pressure-lowering regimen on the risk of recurrent stroke according to stroke subtype and medical history: the PROGRESS Trial. *Stroke* 2004;**35**:116–21.
2. **Lip GY**, Frison L, Grind M. Effect of hypertension on anticoagulated patients with atrial fibrillation. *Eur Heart J* 2007;**28**:752–9.
3. **Rose AJ**, Berlowitz DR, Manze M, *et al*. Comparing methods of measuring treatment intensification in hypertension care. *Circ Cardiovasc Qual Outcomes* 2009;**2**:385–91.
4. **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**:239–45.

Journal of Thrombosis and Haemostasis, **8**: 2182–2191 DOI: 10.1111/j.1538-7836.2010.03996.x

ORIGINAL ARTICLE

Patient characteristics associated with oral anticoagulation control: results of the Veterans AffaiRs Study to Improve Anticoagulation (VARIA)

A. J. ROSE,*† E. M. HYLEK,† A. OZONOFF,*‡, A. S. ASH,†§, J. I. REISMAN* and D. R. BERLOWITZ*†¶
*Center for Health Quality, Outcomes, and Economic Research, Bedford VA Medical Center, Bedford; †Department of Medicine, Section of General Internal Medicine, Boston University School of Medicine, Boston; ‡Department of Biostatistics, Boston University School of Public Health, Boston; §Department of Quantitative Health Sciences, Division of Biostatistics and Health Services Research, University of Massachusetts School of Medicine, Worcester; and ¶Department of Health Policy and Management, Boston University School of Public Health, Boston, MA, USA

To cite this article: 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**: 2182–91.

Summary. *Background:* In patients receiving oral anticoagulation, improved control can reduce adverse outcomes such as stroke and major hemorrhage. However, little is known about patient-level predictors of anticoagulation control. *Objectives:* To identify patient-level predictors of oral anticoagulation control in the outpatient setting. *Patients/Methods:* We studied 124 619 patients who received oral anticoagulation from the Veterans Health Administration from October 2006 to September 2008. The outcome was anticoagulation control, summarized using percentage of time in therapeutic International Normalized Ratio range (TTR). Data were divided into inception (first 6 months of therapy; 39 447 patients) and experienced (any time thereafter; 104 505 patients). Patient-level predictors of TTR were examined by multivariable regression. *Results:* Mean TTRs were 48% for inception management and 61% for experienced management. During inception, important predictors of TTR included hospitalizations (the expected TTR was 7.3% lower for those with two or more hospitalizations than for the non-hospitalized), receipt of more medications (16 or more medications predicted a 4.3% lower than for patients with 0–7 medications), alcohol abuse (− 4.6%), cancer (− 3.1%), and bipolar disorder (− 2.9%). During the experienced period, important predictors of TTR included hospitalizations (four or more hospitalizations predicted 9.4% lower TTR), more medications (16 or more medications predicted 5.1% lower TTR), alcohol abuse (− 5.4%), female sex (− 2.9%), cancer (− 2.7%), dementia (− 2.6%), non-alcohol substance abuse (− 2.4%), and chronic liver disease (− 2.3%). *Conclusions:* Some patients receiving oral anticoagulation therapy are more challenging to maintain within the therapeutic range than others. Our findings can be used to identify patients who require closer attention or innovative management strategies to maximize benefit and minimize harm from oral anticoagulation therapy.

Correspondence: Adam J. Rose, Center for Health Quality, Outcomes, and Economic Research, Bedford VA Medical Center, 200 Springs Road, Building 70, Bedford, MA 01730, USA.
Tel.: +781 687 3198; fax: +917 591 3104.
E-mail: adamrose@bu.edu

Received 1 April 2010, accepted 19 July 2010

Keywords: ambulatory care, anticoagulants, chronic disease, quality of health care, warfarin.

Background

Warfarin is received by millions of patients in the USA for the treatment and prevention of thromboembolic diseases [1]. Although it is highly effective, warfarin has a narrow therapeutic window, and undergoes numerous interactions with drugs, diet, and comorbid conditions [1]. The potential consequences of deviating from the therapeutic range are serious: insufficiently anticoagulated patients remain at risk for thromboembolic events [2], and over-anticoagulated patients are at risk for major hemorrhage, particularly intracranial hemorrhage [3].

Despite the importance of good control for patients receiving warfarin, little is known about patient-level predictors of percentage time in therapeutic International Normalized Ratio (INR) range (TTR) [4], a measure of anticoagulation control over time. White *et al*. [5] identified several patient-level factors that predicted poor control (TTR < 60%): non-white race, paroxysmal as opposed to continuous atrial fibrillation, and being new to warfarin ('inception'). Our group [6] found additional factors that predict poor control: non-standard target INR range (i.e. other than 2.0–3.0), and female sex. These studies examined a relatively narrow range of patient-level factors, and lacked the power to examine uncommon predictors or factors with real, but modest, effects on TTR.

Here, we used data from almost 125 000 patients receiving oral anticoagulation from the Veterans Health Administration

[Veterans AffaiRs Study to Improve Anticoagulation (VARIA)], to explore a more comprehensive range of patient characteristics as potential predictors of anticoagulation control. Our findings could help clinicians and healthcare systems to identify patients who require extra attention or innovative management strategies to achieve acceptable levels of anticoagulation control.

Methods

Data

The VARIA database included all patients deemed to be receiving oral anticoagulation therapy (OAT) from the Veterans Health Administration (VA) between 1 October 2006 and 30 September 2008. The VA, the largest integrated health system in the USA, delivered care to 5.6 million patients in 2009, representing 1.8% of the US population [7]. Patients become eligible to receive VA care by serving in the military for two or more years; eligibility is for life. Many patients receive all of their care in the VA system, although some also visit non-VA healthcare facilities for a portion of their care. The VA collects comprehensive data regarding the care delivered within the VA system, including demographics, dates of service, pharmacy, and laboratory data. Laboratory data, including INR values, are first entered into local clinical databases (uploaded from the laboratory analyzer), and then these local databases are periodically uploaded to the national VA Corporate Data Warehouse, the source for our data. The study was approved by the Institutional Review Board of the Bedford VA Medical Center.

Each VA site consists of a central hospital along with its on-site outpatient clinics and several outlying satellite clinics. By policy, all VA anticoagulation care is delivered in dedicated anticoagulation clinics (ACCs) [8]. However, VA ACCs differ in organization, management, and performance. We address these between-site differences in a separate article.

We included INR tests within the VA when patients were 'on warfarin': that is, when a patient was either (i) 'in possession' of warfarin or (ii) having INR tests every 42 days. A similar approach was used to define time 'on warfarin' in a previous study [9]. 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 more than 30 days beyond the end of a prescription does not necessarily indicate that warfarin therapy has stopped. We therefore also allowed a consistent pattern of INR measurements (i.e. every 42 days or less) to indicate that a patient was still being managed. Patients with chronic liver disease only qualified though receipt of warfarin; frequent INR tests for such a patient might have been performed to monitor liver function. We excluded INR tests performed while the patient was hospitalized within the VA system. Patients who are hospitalized may receive temporary parenteral anticoagulation (e.g. with heparin) or no anticoagulation, so low INR values are likely to be intentional.

Dependent variable: TTR

We calculated TTR (between 0% and 100%) using Rosendaal's method [4], which uses linear interpolation to assign an INR value to each day between two 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 and 3 is calculated [4]. Previous studies have linked TTR and clinical outcomes such as stroke, venous thromboembolism (VTE), and major hemorrhage [5,10], thus validating this intermediate outcome measure for use as a surrogate endpoint for OAT [11].

Independent variable: primary indication for OAT

Patients receiving OAT for different indications may vary in duration of therapy, INR target range, health status, and other factors. We developed an algorithm to assign a primary indication for therapy to patients with more than one of these conditions. A list of the ICD-9 codes used to define these conditions is given in Appendix S1. When present, valvular heart disease (VHD) was considered to be the primary indication; if VHD was absent, VTE was the primary indication; if both VHD and VTE were absent, atrial fibrillation (AF) was the primary indication. If VHD, VTE and AF were absent, a group called 'other indications' (including cardiomyopathy, left ventricular thrombus, and left ventricular aneurysm) encompassed all other therapeutic indications [1].

Independent variables: demographics

We collected patient demographics, including sex, age, race/ethnicity, and zip code of residence. Race/ethnicity was classified as non-Hispanic White, non-Hispanic Black, Hispanic, Asian, Native American, and Missing (11%). We did not attempt to impute race, because of concern that it was not missing by chance [12]. The patient's zip code of residence was linked to US Census data [13] to obtain the percentage of people living below the federal poverty line in each geographic area, which was used as a proxy measure for socio-economic status [14]. We also used the patient's zip code of residence to obtain the straight-line driving distance between the centroid representation of the zip code and the nearest VA healthcare facility.

Independent variable: warfarin experience

We have previously observed that anticoagulation control differs markedly between the first 6 months of therapy ('inception') and the period thereafter ('experienced') [6]. We therefore examined predictors of TTR separately for each period. We defined each patient's date of warfarin initiation, looking back as far as 1 October 2001. Initiation was defined as the first INR value greater than 1.2 or the first outpatient warfarin fill, whichever came first. It would be extremely unusual for a patient to record an INR value above 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). A single patient might contribute only to the inception dataset (if he or she had less than 6 months of therapy), only to the experienced dataset (if he or she began warfarin more than 6 months prior to the inception of our study), or to both. Having controlled for inception data through stratification, we also considered duration of warfarin experience as a potential indicator within the experienced dataset.

Independent variables: clinical variables

We used ICD-9 codes to define comorbid conditions as detailed in Appendix S2. Previous studies have demonstrated that ICD-9 codes can be used successfully to identify comorbid conditions within VA databases [15,16]. Conditions were examined because they were expected to affect anticoagulation control. Examples include conditions already shown to worsen anticoagulation control (e.g. cancer) [17], conditions treated with medications that interact with warfarin (e.g. epilepsy and bipolar disorder), conditions associated with chaotic lifestyle or poor adherence (e.g. dementia and substance abuse), and conditions that directly interfere with hepatic function (e.g. alcohol abuse and chronic liver disease). To further characterize each patient's illness burden, we also counted the number of times each patient was hospitalized within the VA system during the inception period (zero, one, or two or more) and during the experienced period (zero, one, two, three, or four or more), as well as the total number of distinct classes of non-warfarin medications received chronically (for at least 30 days) during the study period (0–7, 8–11, 12–15, and ≥ 16).

Assembling the final database

Figure 1 describes our database construction. First, we required at least two interpolable between-test intervals of 56 days or fewer for study entry, to ensure that our population consisted of 'VA users', or patients who use the VA for most or all of their care. We excluded patients whose primary indication for warfarin was VHD, since their target INR range might be 2.5–3.5 rather than 2–3. We also excluded patients who only recorded INR values 1.2 and lower, reasoning that their INR tests were unlikely to be related to warfarin management. Finally, we required at least two interpolable between-test intervals of 56 days or fewer for study entry, to ensure that our population consisted of 'VA users', or patients who use the VA for most or all of their care.

There are 128 sites of care in the VA system. We excluded 28 of these sites from our study and also excluded 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. Whereas accurate data are collected about which laboratory tests are performed (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

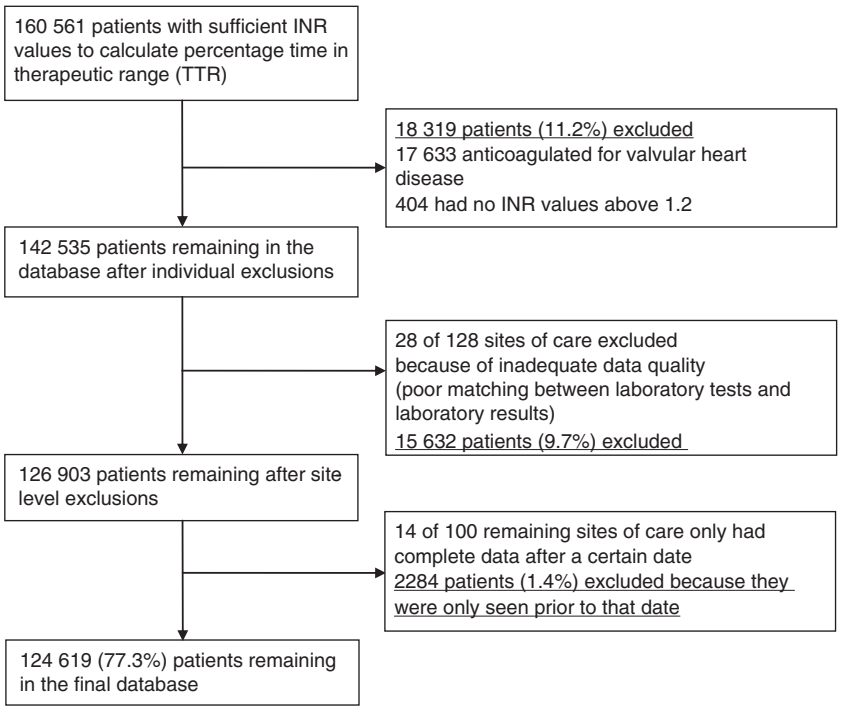


Fig. 1. Enrollment flowchart. INR, International Normalized Ratio.

several months when 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 performed 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 and continued to have it through to the end. Thus, 28 sites were excluded because of incomplete data, and 14 sites were partially included.

Statistical analyses

We calculated TTR for each patient; for patients with both inception and experienced data, we calculated TTR separately for each period. We also analyzed the inception and experienced data separately, in each case first examining the effect of each variable upon TTR individually and then in a fully adjusted multivariable model. We used linear regression for our adjusted models, employing a mixed model (SAS PROC MIXED) with exchangeable correlation structure to account for the correlation of patient outcomes by site of care. We performed all analyses with SAS version 9.1 (SAS Corporation, Cary, NC, USA).

Creation of a clinical prediction tool

When considering the initiation of warfarin therapy, clinicians may find it useful to know which patients could be difficult to control in the future. We therefore developed a clinical prediction tool to help predict a patient's future TTR at the time of initiation of therapy. The predicted TTR during the experienced period may be especially helpful, so the prediction tool provides separate estimates for the first 6 months of therapy (inception) and the 18 months after that (experienced). The models underlying the tool were generally based on our main results, but we re-derived the models using only data that would be available to clinicians at the time of initiation.

To enable our model to predict TTR during the inception period, we located all patients in our database who were new to warfarin and remained on it for a full 6 months thereafter (n = 25 788). To enable our model to predict TTR during the experienced period, we located all patients who were managed in the experienced period for at least 18 months (n = 86 731). We assessed all the predictor variables at baseline, that is, during the year prior to the patient's first INR value in our model. We divided each of these datasets into a larger (60%) derivation set and a smaller (40%) validation set. We derived and validated a somewhat simplified model to predict the mean TTR over the first 6 months (inception) and the succeeding 18 months (experienced). The model was simplified as follows: race was removed (because of our concern that keeping race in the model might perpetuate racial disparities in care), poverty and distance were removed (their effects were small, and clinicians might not be able to assess them easily), and the list of comorbid conditions was abbreviated to those that had statistically significant effects in at least one time period. In addition, only hospitalizations that occurred during the year prior to warfarin initiation were considered as a predictor of TTR. On the basis of the parameters of this simplified model, we then created a clinical prediction tool in the form of a spreadsheet.

Results

Enrollment and baseline characteristics

After exclusion of individual patients and sites of care with missing data, there were 124 619 patients who received anticoagulation from 100 sites of care. There were 163 144 total patient-years of observation in the database. TTR could not be calculated during 34 963 of these 163 144 patient-years (21.4% of all patient-time), because of hospitalizations or gaps in therapy.

Details of the study sample are given in Table 1 (inception) and Table 2 (experienced management). The following results are for patient characteristics during the inception period; experienced period results were similar. The sample was overwhelmingly male. The age distribution reflects the fact that the indications for anticoagulation are most common in the elderly. A considerable number of minority patients were enrolled (e.g. 11.3% non-Hispanic Black). Most patients were anticoagulated for AF (55%), with the remainder being anticoagulated for VTE (35%) or other indications (10%). Patients had a considerable burden of comorbid illness. For example, 80% had hypertension, 37% had diabetes, 26% had heart failure, and 10% received a new diagnosis of cancer during the study. Mental health and substance abuse conditions were also common: 24% had major depression, 14% abused alcohol, and 7% abused some other substance. Medication and hospitalization data also reflect high levels of comorbid illness: 61% received at least eight non-warfarin chronic medications, and 22% were hospitalized at least once during the 6-month inception period.

Predictors of TTR: inception period

We observed inception (first 6 months of therapy) for 39 447 patients (Table 1). Mean TTR during inception was 48%. In the adjusted analysis, younger age predicted worse control; patients under age 55 years had TTR 3.9% lower than the reference category (age ≥ 75 years). Most racial minorities had lower TTR during inception, although these relationships were attenuated after multivariable adjustment. Poverty in the zip code of residence predicted worse control, with residents of the poorest areas having TTR 2.7% lower than the wealthiest. Driving distance to the nearest VA had a small negative influence upon inception-period TTR, but only when the distance was 20 miles or greater (− 1.3%). Most comorbid conditions reduced inception-period TTR. Among the physical illnesses that had the strongest adverse effects during inception

Table 1 Patient characteristics and effects on percentage time in therapeutic International Normalized Ratio range during the inception period, that is, the first 6 months of warfarin therapy (n = 39 447)

Variable	Number (%)	Unadjusted effect (95% CI)	Adjusted effect (95% CI)
Intercept			52.4
Female sex	1046 (2.7)	−0.6 (−1.3 to 0.1)	+0.6 (−1.0 to 2.2)
Age group (years)			
20–54	4993 (12.7)	−6.0 (−6.4 to −5.6)**	−3.9 (−4.9 to −2.9)**
55–59	6404 (16.2)	−3.3 (−3.6 to −2.9)**	−1.7 (−2.5 to −0.8)**
60–64	6599 (16.7)	−1.7 (−2.0 to −1.3)**	−0.8 (−1.6 to 0.1)
65–69	4335 (11.0)	−1.2 (−1.6 to −0.8)**	−0.6 (−1.6 to 0.3)
70–74	5404 (13.7)	−0.7 (−1.1 to −0.4)**	−0.3 (−1.2 to 0.5)
≥ 75	11 712 (29.7)	–	–
Race/ethnicity			
Non-Hispanic White	29 137 (73.9)	–	–
Non-Hispanic Black	4459 (11.3)	−5.6 (−5.9 to −5.2)**	−2.9 (−3.8 to −2.0)**
Hispanic	1149 (2.9)	−4.2 (−4.9 to −3.5)**	−2.7 (−4.3 to −1.1)*
Asian	134 (0.3)	−1.1 (−3.1 to 0.9)**	−1.7 (−6.2 to 2.7)
Native American	148 (0.4)	+2.8 (0.9 to 4.7)**	+3.8 (−0.4 to 8.0)
Other/unknown	4420 (11.2)	+3.1 (2.7 to 3.5)**	+2.0 (1.2 to 2.9)**
Percentage poverty in zip code of residence (quintile)			
Wealthiest (0–5.9)	7697 (19.5)	–	–
Wealthy (5.9–9.0)	7865 (19.9)	−1.1 (−1.5 to −0.7)**	−0.5 (−1.3 to 0.3)
Moderate (9.0–12.6)	7753 (19.7)	−1.7 (−2.1 to −1.4)**	−0.9 (−1.8 to −0.1)*
Poor (12.6–17.8)	7724 (19.6)	−2.8 (−3.2 to −2.4)**	−1.5 (−2.3 to −0.6)*
Poorest (17.8–100)	8408 (21.3)	−5.3 (−5.7 to −5.0)**	−2.7 (−3.6 to −1.9)**
Driving distance from nearest VA facility in miles (quintile)			
Nearest (0–3.1)	8022 (20.3)	–	–
Near (3.1–6.0)	8167 (20.7)	+0.5 (0.2 to 0.9)**	+0.2 (−0.6 to 1.0)
Moderate (6.0–10.5)	8110 (20.6)	+1.5 (1.2 to 1.9)**	+0.4 (−0.4 to 1.2)
Far (10.5–20.3)	7811 (19.8)	+1.3 (0.9 to 1.6)**	+0.1 (−0.8 to 0.9)
Furthest (>20.3)	7337 (18.6)	−0.5 (−0.9 to −0.1)**	−1.3 (−2.1 to −0.4)*
Primary indication for warfarin†			
Atrial fibrillation	21 584 (54.7)	–	–
Venous thromboembolism	13 951 (35.4)	−0.8 (−1.1 to −0.6)**	+1.4 (0.8 to 2.0)**
All others combined	3912 (9.9)	−1.9 (−2.3 to −1.5)**	−1.1 (−2.0 to −0.2)*
Physical comorbid conditions			
Cancer (newly diagnosed)	3945 (10.0)	−3.5 (−3.9 to −3.1)**	−3.1 (−3.9 to −2.2)**
Chronic kidney disease	5233 (13.3)	−2.8 (−3.2 to −2.5)**	−0.9 (−1.7 to −0.1)*
Chronic liver disease	565 (1.4)	−5.7 (−6.6 to −4.7)**	−0.8 (−3.0 to 1.4)
Chronic lung disease	11 018 (27.9)	−2.3 (−2.6 to −2.0)**	−0.1 (−0.8 to 0.5)
Coronary artery disease	16 654 (42.2)	−0.7 (−0.9 to −0.4)**	−0.1 (−0.7 to 0.5)
Diabetes	14 433 (36.6)	−2.0 (−2.2 to −1.7)**	−1.3 (−1.9 to −0.7)**
Epilepsy	1091 (2.8)	−3.6 (−4.3 to −2.9)**	−1.0 (−2.6 to 0.5)
Heart failure	10 139 (25.7)	−2.6 (−2.9 to −2.4)**	−0.3 (−1.0 to 0.3)
Hyperlipidemia	26 983 (68.4)	+2.1 (1.8 to 2.3)**	+2.5 (1.9 to 3.1)**
Hypertension	31 368 (79.5)	−0.4 (−0.7 to −0.1)*	+0.5 (−0.2 to 1.2)
Pain disorders	28 281 (71.7)	−2.3 (−2.6 to −2.1)**	−0.3 (−0.9 to 0.3)
Peripheral arterial disease	6390 (16.2)	−1.7 (−2.0 to −1.3)**	−0.6 (−1.3 to 0.1)
Mental comorbid conditions			
Alcohol abuse	5345 (13.5)	−7.4 (−7.7 to −7.0)**	−4.6 (−5.4 to −3.7)**
Anxiety	4369 (11.1)	−1.9 (−2.3 to −1.6)**	+0.8 (−0.1 to 1.7)
Bipolar disorder	1279 (3.2)	−6.8 (−7.5 to −6.2)**	−2.9 (−4.5 to −1.4)**
Dementia	1726 (4.4)	−2.1 (−2.7 to −1.6)**	−1.6 (−2.9 to −0.3)*
Major depression	9389 (23.8)	−3.4 (−3.6 to −3.1)**	−0.6 (−1.3 to 0.1)
PTSD	3718 (9.4)	−3.1 (−3.5 to −2.7)**	+0.3 (−0.7 to 1.2)
Schizophrenia	701 (1.8)	−5.9 (−6.8 to −5.0)**	−0.7 (−2.7 to 1.3)
Substance abuse (non-alcohol)	2677 (6.8)	−8.7 (−9.2 to −8.2)**	−2.4 (−3.5 to −1.2)**
Number of non-warfarin medications			
0–7	15 416 (39.1)	–	–
8–11	12 231 (31.0)	−2.5 (−2.8 to −2.3)**	−1.7 (−2.4 to −1.0)**
12–15	7266 (18.4)	−5.4 (−5.7 to −5.0)**	−3.4 (−4.2 to −2.5)**
≥ 16	4534 (11.5)	−7.6 (−8.0 to −7.2)**	−4.3 (−5.4 to −3.3)**
Number of hospitalizations during inception period			
None	30 675 (77.8)	–	–
1	5860 (14.9)	−4.9 (−5.2 to −4.6)**	−3.1 (−3.9 to −2.4)**
≥ 2	2912 (7.4)	−10.1 (−10.6, −9.7)**	−7.3 (−8.3, −6.2)**

CI, confidence interval; PTSD, post-traumatic stress disorder; VA, Veterans Health Administration. All β-coefficients are in units of percentage time in therapeutic International Normalized Ratio range. All P-values account for the correlation of outcomes by site of care. Adjusted effects are adjusted for all the other variables in the table.

*$P < 0.05$.

**$P < 0.001$.

†Patients whose main indication for anticoagulation was valvular heart disease or prosthetic heart valve were excluded from this study.

Table 2 Patient characteristics and effects on percentage time in therapeutic International Normalized Ratio range during the experienced period, that is, any time after the first 6 months of warfarin therapy (n = 104 505)

Variable	Number (%)	Unadjusted effect (95 CI)	Adjusted effect (95% CI)
Intercept			63.2
Female sex	1984 (1.9)	−5.5 (−5.9 to −5.0)**	−2.9 (−3.9 to −2.0)**
Age group (years)			
20–54	7430 (7.1)	−9.2 (−9.5 to −9.0)**	−4.7 (−5.3 to −4.1)**
55–59	11 590 (11.1)	−4.8 (−5.0 to −4.6)**	−1.2 (−1.7 to −0.7)**
60–64	12 783 (12.2)	−2.5 (−2.6 to −2.3)**	+0.1 (−0.3 to 0.5)
65–69	11 705 (11.2)	−0.1 (−0.3 to 0.1)	+1.0 (0.6 to 1.5)**
70–74	17 046 (16.3)	+0.5 (0.3 to 0.7)**	+1.1 (0.7 to 1.4)**
≥ 75	43 951 (42.1)	–	–
Race/ethnicity			
Non-Hispanic White	80 728 (77.2)	–	–
Non-Hispanic Black	8853 (8.5)	−6.5 (−6.7 to −6.3)**	−2.6 (−3.1 to −2.1)**
Hispanic	2977 (2.8)	−2.0 (−2.3 to −1.6)**	−0.6 (−1.4 to 0.2)
Asian	302 (0.3)	−0.3 (−1.4 to 0.8)	−0.5 (−2.9 to 1.9)
Native American	279 (0.3)	−4.3 (−5.5 to −3.2)**	−2.4 (−4.9 to 0.1)
Other/unknown	11 366 (10.9)	+2.4 (2.3 to 2.6)**	+0.3 (−0.2 to 0.7)
Percentage poverty (quintiles)			
Wealthiest (0.0–5.9)	21 193 (20.3)	–	–
Wealthy (5.9–9.0)	20 933 (20.0)	−0.5 (−0.7 to −0.3)**	+0.1 (−0.3 to 0.5)
Moderate (9.0–12.6)	20 987 (20.1)	−1.4 (−1.6 to −1.2)**	−0.5 (−0.9 to −0.1)*
Poor (12.6–17.8)	20 910 (20.0)	−2.0 (−2.2 to −1.8)**	−0.8 (−1.3 to −0.4)**
Poorest (17.8–100.0)	20 482 (19.6)	−4.0 (−4.2 to −3.8)**	−1.5 (−2.0 to −1.1)**
Driving distance to nearest VA in miles (quintiles)			
Nearest (3.1 or closer)	21 119 (20.2)	–	–
Near (3.1–6.0)	20 985 (20.1)	+0.4 (0.2 to 0.6)**	+0.1 (−0.3 to 0.5)
Moderate (6.0–10.5)	21 094 (20.2)	+0.5 (0.3 to 0.7)**	−0.4 (−0.8 to 0.0)
Far (10.5–20.3)	21 024 (20.1)	+1.0 (0.8 to 1.2)**	+0.0 (−0.4 to 0.4)
Furthest (20.3 or farther)	20 283 (19.4)	+0.5 (0.3 to 0.7)**	+0.0 (−0.4 to 0.4)
Date of warfarin inception			
≥6 months before study inception	78 142 (74.8)	–	–
6 months before the study, first year of the study	8885 (8.5)	−4.5 (−4.7 to −4.4)**	−3.6 (−3.9 to −3.3)**
Second year of the study	3658 (3.5)	−6.0 (−6.3 to −5.6)**	−5.8 (−6.5 to −5.1)**
Primary indication for warfarin[†]			
Atrial fibrillation	67 077 (64.2)	–	–
Venous thromboembolism	28 585 (27.4)	−3.8 (−4.0 to −3.7)**	−1.2 (−1.5 to −0.9)**
All others combined	8843 (8.5)	−1.5 (−1.7 to −1.3)**	−1.5 (−1.9 to −1.0)**
Physical comorbid conditions			
Cancer (newly diagnosed)	7100 (6.8)	−4.8 (−5.0 to −4.5)**	−2.7 (−3.2 to −2.2)**
Chronic kidney disease	14 806 (14.2)	−4.4 (−4.6 to −4.3)**	−1.6 (−2.0 to −1.2)**
Chronic liver disease	1253 (1.2)	−8.3 (−8.8 to −7.7)**	−2.3 (−3.5 to −1.1)**
Chronic lung disease	30 687 (29.4)	−3.8 (−3.9 to −3.7)**	−0.7 (−1.0 to −0.4)**
Coronary artery disease	53 114 (50.8)	−1.4 (−1.5 to −1.2)**	−0.6 (−0.9 to −0.3)**
Diabetes	41 863 (40.1)	−2.1 (−2.2 to −2.0)**	−1.0 (−1.3 to −0.7)**
Epilepsy	2926 (2.8)	−5.1 (−5.4 to −4.7)**	−1.6 (−2.4 to −0.8)**
Heart failure	34 229 (32.8)	−3.6 (−3.7 to −3.5)**	−1.0 (−1.3 to −0.7)**
Hyperlipidemia	78 754 (75.4)	+2.1 (2.0 to 2.2)**	+2.0 (1.7 to 2.3)**
Hypertension	87 776 (84.0)	+0.0 (−0.2 to 0.1)	+1.0 (0.7 to 1.4)**
Pain disorders	76 159 (72.9)	−3.4 (−3.5 to −3.3)**	−0.3 (−0.6 to 0.0)
Peripheral arterial disease	20 746 (19.9)	−2.3 (−2.5 to −2.2)**	−0.5 (−0.8 to −0.1)*
Mental comorbid conditions			
Alcohol abuse	9729 (9.3)	−9.4 (−9.6 to −9.2)**	−5.4 (−5.9 to −4.9)**
Anxiety	10 253 (9.8)	−4.3 (−4.5 to −4.1)**	−0.2 (−0.6 to 0.3)
Bipolar disorder	2386 (2.3)	−8.4 (−8.8 to −8.0)**	−1.8 (−2.7 to −1.0)**
Dementia	5517 (5.3)	−4.2 (−4.4 to −3.9)**	−2.6 (−3.2 to −2.0)**
Major depression	22 583 (21.6)	−5.9 (−6.1 to −5.8)**	−2.0 (−2.3 to −1.6)**
PTSD	8066 (7.7)	−5.0 (−5.2 to −4.7)**	+0.4 (−0.2 to 0.9)
Schizophrenia	1263 (1.2)	−6.7 (−7.2 to −6.1)**	+0.8 (−0.4 to 2.0)
Substance abuse (non-alcohol)	4233 (4.1)	−11.8 (−12.1 to −11.5)**	−2.4 (−3.2 −1.7)**
Number of non-warfarin medications			
0–7	43 380 (41.5)	–	–
8–11	33 393 (32.0)	−3.6 (−3.7 to −3.4)**	−1.8 (−2.1 to −1.5)**
12–15	17 915 (17.1)	−7.3 (−7.4 to −7.1)**	−3.2 (−3.6 to −2.8)**

Table 2 (Continued)

Variable	Number (%)	Unadjusted effect (95 CI)	Adjusted effect (95% CI)
≥ 16	9817 (9.4)	−11.7 (−11.9 to −11.5)**	−5.1 (−5.6 to −4.5)**
Number of hospitalizations during experienced period			
None	77 107 (73.8)	–	–
1	13 858 (13.3)	−6.2 (−6.4 to −6.0)**	−3.7 (−4.1 to −3.3)**
2	6261 (6.0)	−8.5 (−8.8 to −8.3)**	−5.1 (−5.7 to −4.5)**
3	3066 (2.9)	−10.6 (−11.0 to −10.3)**	−6.5 (−7.3 to −5.7)**
≥ 4	4213 (4.0)	−14.9 (−15.2 to −14.6)**	−9.4 (−10.1 to −8.7)**

CI, confidence interval; PTSD, post-traumatic stress disorder; VA, Veterans Health Administration. All β-coefficients are in units of percentage time in therapeutic International Normalized Ratio range. All P-values account for the correlation of outcomes by site of care. Adjusted effects are adjusted for all the other variables in the table.

*$P < 0.05$.

**$P < 0.001$.

†Patients whose main indication for anticoagulation was valvular heart disease or prosthetic heart valve were excluded from this study.

were cancer (− 3.1%), diabetes (− 1.3%), and chronic kidney disease (− 0.9%). In general, mental illnesses exerted a stronger effect than physical illnesses, especially alcohol abuse (− 4.6%), bipolar disorder (− 2.9%), substance abuse (− 2.4%), and dementia (− 1.6%). There were several conditions that had paradoxical effects (associated with improved control), most notably hyperlipidemia (+ 2.5%). The number of chronic non-warfarin medications was inversely related to inception TTR: those receiving the most medications had TTR 4.3% lower than those receiving the least. Although inpatient INR values were excluded from calculations of TTR, hospitalizations were still associated with lower TTR, with those hospitalized two or more times having TTR 7.3% lower than those not hospitalized.

Predictors of TTR: experienced period

We observed post-inception ('experienced period') therapy for 104 505 patients (Table 2). Mean TTR during the experienced period was 61%. In the adjusted analysis, women had lower TTR (− 2.9%), an effect not seen during inception. Similar to what was found for the inception period, age less than 55 years predicted lower TTR (− 4.7%), and Black patients had lower TTR (− 2.6%) than Whites. Area poverty predicted lower TTR, but the effect size in the experienced period (1.5% lower in the poorest areas) was smaller than it had been during inception. In contrast to what was found for the inception period, driving distance was unrelated to TTR. Even during the experienced period, further experience with warfarin was associated with improved control: TTR was highest among patients who began warfarin at least 6 months before the study began.

As in the inception period, most comorbid conditions predicted lower TTR. Medical conditions with the strongest adverse effects included cancer (− 2.7%), chronic liver disease (− 2.3%), epilepsy (− 1.6%), chronic kidney disease (− 1.6%), diabetes (− 1.0%), and heart failure (− 1.0%). Mental health conditions had stronger effects, particularly alcohol abuse (− 5.4%), dementia (− 2.6%), substance abuse (− 2.4%), major depression (− 2.0%), and bipolar disorder (− 1.8%). As was seen during inception, several conditions were somewhat surprisingly associated with improved control, especially hyperlipidemia (+ 2.0%). Medications and hospitalizations were powerful predictors of control, even after adjustment for comorbid conditions. For example, patients hospitalized four or more times had TTR 9.4% lower than those not hospitalized.

Clinical prediction tool

We developed a tool to help clinicians to predict inception and experienced TTR for each patient on the basis of patient-level characteristics available at the time of initiation of therapy (Appendix S3). The tool is available as an online-only supplement to this article, and may be downloaded and used to predict TTR on the basis of patient characteristics known at the time of inception. The model parameters were generally quite similar to those presented in Tables 1 and 2. R^2 for inception TTR was 3.4% in the derivation set and 3.2% in the validation set. R^2 for experienced TTR was 6.5% in the derivation set and 6.8% in the validation set.

Discussion

Using data from the VA, we examined patient-level factors related to anticoagulation control with warfarin, measured asTTR. Our study demonstrates that some patients are more difficult to keep within the target INR range than others. We estimated the effects of multiple patient-level characteristics on TTR, clarifying the contributions of illness burden and clinical complexity to poorer anticoagulation control. Alcohol abuse, substance abuse, cancer and dementia were particularly strong predictors of poor control. The persistence of hospitalizations as a strong predictor of TTR, even after controlling for a broad range of comorbid conditions, bears mention. Because we did not consider inpatient INR values, this finding suggests that the period after a hospitalization is also characterized by poorly controlled anticoagulation. A hospitalization event entails numerous changes in diet, lifestyle, and health status, which can perturb the management of warfarin, as has been noted in earlier studies [18]. Patients who have been hospitalized require prompt and

vigilant follow-up after discharge to prevent such derangements, which have now been seen in multiple studies spanning multiple settings.

Several conditions were associated with improved TTR, most notably hyperlipidemia. To explore this, we compared patients receiving HMG-CoA reductase inhibitors (statins) for hyperlipidemia to those receiving fibrates, hypothesizing that statins themselves may improve TTR; however, the groups did not differ. Another possibility is that the presence of a code for hyperlipidemia (an asymptomatic risk factor) is a proxy for a proactive patient–clinician health maintenance partnership. This is supported by the similar positive effect for hypertension.

We found that achieving good control with warfarin is harder during the inception period than during the experienced period (mean TTR 48% vs. 61%), and that predictors of TTR differ between the two periods. Our study also suggests that, even within the VA, a system that provides comprehensive coverage and access, disparities in anticoagulation control still exist, with poorer, more distant, female and minority patients all experiencing poorer control in one or both of our study cohorts. In a previous non-VA study, our group also found that women have lower TTR than men [6]. The reason for this finding is unclear, but it is apparently not limited to the VA, and should be investigated further. In addition, younger patients experienced worse TTR. Although we cannot be certain, it is possible that the demands of a full-time job may be difficult to reconcile with the frequent follow-up expected of anticoagulation patients. On the basis of this finding, the VA may wish to consider improving access to care for working patients receiving anticoagulation.

Previous studies have explored only a few possible predictors of TTR, including inception status, race, and sex [5,6]. Other studies have examined the effect of a single predictor in detail, including cancer [17], non-white race [19–21], health literacy [22], and the patient's understanding of the reason why warfarin was prescribed [23]. Although our study echoes some of these earlier findings, we examined a much more comprehensive set of predictors and adjusted for confounding, strengths not found in previous studies. In addition, the size and scope of our database allowed us to examine relatively uncommon predictors such as chronic liver disease and epilepsy.

The primary purpose of the models presented here is risk adjustment. Risk adjustment allows fair comparisons of outcomes between sites of care, despite the fact that some sites will have more challenging patient populations than others [24]. Thus, these models may be used to compare sites of care on risk-adjusted TTR, which could spur quality improvement and improve patient outcomes [11]. The results reported here will underpin a planned effort to improve anticoagulation control for VA patients receiving warfarin.

Our findings can also be used to identify patients who may require greater attention or innovative management strategies to achieve acceptable levels of anticoagulation control. The effect of each individual variable upon TTR in our study was relatively small, and no single risk factor that we studied establishes a patient as being too high-risk for anticoagulation therapy to be contemplated. In combination, however, the variables that we studied would indeed be useful for risk-stratifying patients receiving oral anticoagulation. To facilitate the use of this information to guide clinical practice, we have offered, with this article, a simplified version of our models in the form of a clinical prediction tool. This tool can help clinicians to predict a patient's eventual level of anticoagulation control before starting therapy. Such information could be helpful in determining which patients might require aggressive interventions to help them achieve an acceptable level of control, and which patients have such poor predicted control that they might not be suitable candidates for warfarin. Although warfarin is a superior therapy, aspirin is available as a second-line agent, at least for patients with AF [25], and its safety and effectiveness are unlikely to be compromised by patient characteristics in the same way.

In addition, it seems likely that novel oral anticoagulants may soon be approved for use, particularly dabigatran [26,27]. On the basis of studies of dabigatran and previous studies of other similar drugs, novel anticoagulants are likely to offer similar safety and effectiveness to warfarin, at a much higher cost [28], but might be a superior choice for patients with poor control on warfarin (although this has not yet been proven). Models such as the ones presented here could possibly be used prospectively to identify some patients who are likely to do well with warfarin and others who are unlikely to do well and might be candidates for the novel agents. Restricting the use of these agents to patients who are poor candidates for warfarin could markedly improve their cost-effectiveness.

Although our study is unprecedented in the amount and richness of the data collected to study anticoagulation, we do acknowledge some important limitations. First, there are some caveats regarding the clinical prediction tool. This tool is intended to predict TTR at the time of initiation; after a patient has received warfarin for several months, his or her past control will clearly be the best predictor of future control. In addition, a clinical prediction tool should not only be validated in the source population ('internal validation'), as ours was, but should also be validated again in a separate population ('external validation'). It is even more useful if the investigators can demonstrate the impact of the tool on patient outcomes as compared with usual practice [29,30]. Our rule has not yet been externally validated, and nor has its impact on clinical care been demonstrated. Clinicians should access this tool if they find it useful, but should be aware that it has not yet achieved these benchmarks for the highest-quality clinical prediction tools. In future research, we hope to accomplish these tasks to help this tool gain wider acceptance and increase its impact on clinical management.

Second, VA patients are mostly male and have a high burden of physical and mental health conditions. It is unclear how our results might have differed in a population of patients with more women and a lesser degree of comorbidity. Third, we studied an intermediate outcome of care (TTR) rather than definitive outcomes such as stroke or major hemorrhage.

However, TTR has been convincingly linked to definitive outcomes [5,10], making higher TTR a good target for improving health outcomes [11]. Fourth, our administrative database did not contain information on some known determinants of anticoagulation control, such as adherence to therapy and dietary variation. Such data could only be collected through expensive procedures such as frequent questionnaires, which would greatly limit sample size. This would have precluded a study of such size and scope. In addition, the intended use of our models is to risk-adjust TTR, a task that must be accomplished with administrative data alone to be feasible. Fifth, our database did not contain information on care received outside the VA, and there is no comprehensive database of INR values obtained outside the VA system. This precluded a consideration of INR values obtained outside the VA. However, because we required at least two between-test intervals of 56 days or fewer for study entry, the great majority of the patients that we studied would have been 'VA users', who are likely to visit the VA for most, if not all, of their care.

Finally, although ICD-9 codes have been used in many previous studies to identify comorbid conditions, we acknowledge that this approach may lack sensitivity, particularly for some conditions that are stigmatized or difficult to recognize, such as alcohol abuse or dementia. We addressed this concern in part by requiring only one ICD-9 code for these conditions rather than two, a strategy that increased their frequency somewhat and thus presumably improved sensitivity (see Appendix S2 for details). Through this and other carefully considered decisions, we have tried to improve the performance of ICD-9 codes in identifying comorbid conditions as much as possible.

In conclusion, we have collected a database of patients receiving oral anticoagulation that is unprecedented in its size and scope. We used this database to identify the combined effect of multiple patient-level predictors of anticoagulation control. These data and models could enable the VA to measure and improve the quality of its oral anticoagulation care.

Acknowledgements

The authors thank B. Bokhour for her contributions to this study. The opinions expressed in this manuscript do not necessarily represent the official views of the Department of Veterans Affairs.

Disclosure of Conflict of Interests

A. J. Rose is supported by a Career Development Award from the Veterans Affairs Health Services Research and Development Service. E. M. Hylek has received research funding from Bristol Myers Squibb. She also reports receiving honoraria from Bayer and Bristol Myers Squibb, as well as serving on advisory boards for Boehringer-Ingelheim, Bristol Myers Squibb, and Sanofi Aventis. The other authors state that they have no conflict of interest.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1. Codes to define indications for oral anticoagulation therapy.
Appendix S2. Codes to define comorbid conditions.
Appendix S3. TTR calculator.

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 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–98S.
2 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**: 591–7.
3 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.
4 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**: 236–9.
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**: 239–45.
6 Rose AJ, Ozonoff A, Henault LE, Hylek EM. Warfarin for atrial fibrillation in community-based practice. *J Thromb Haemost* 2008; **6**: 1647–54.
7 United States Department of Veterans Affairs. 2009 Annual Performance and Accountability Report. www4.va.gov/budget/report. Accessed 25 May 2010.
8 United States Department of Veterans Affairs. VHA Directive 2009-003. Anticoagulation Therapy Management. 2 February 2009.
9 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**: 2685–92.
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**: 2029–37.
11 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**: 146–55.
12 Little RJA, Rubin DB. *Statistical Analysis with Missing Data*. Hoboken, NJ: Wiley, 2002.
13 United States Census Bureau. United States Census 2000. http://www.census.gov/main/www/cen2000.html. Accessed 20 November 2009.
14 Krieger N, Waterman PD, Chen JT, Rehkopf DH, Subramanian SV. Geocoding and monitoring US socioeconomic inequalities in health: an introduction to using area-based socioeconomic measures – The

Public Health Disparities Geocoding Project monograph. http://www.hsph.harvard.edu/thegeocodingproject/. Accessed 1 April 2010.
15 Borzecki AM, Wong AT, Hickey EC, Ash AS, Berlowitz DR. Identifying hypertension-related comorbidities from administrative data: what's the optimal approach? *Am J Med Qual* 2004; **19**: 201–6.
16 Szeto HC, Coleman RK, Gholami P, Hoffman BB, Goldstein MK. Accuracy of computerized outpatient diagnoses in a Veterans Affairs general medicine clinic. *Am J Manag Care* 2002; **8**: 37–43.
17 Rose AJ, Sharman JP, Ozonoff A, Henault LE, Hylek EM. Effectiveness of warfarin among patients with cancer. *J Gen Intern Med* 2007; **22**: 997–1002.
18 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**: 705–14.
19 Bhandari VK, Wang F, Bindman AB, Schillinger D. Quality of anticoagulation control: do race and language matter? *J Health Care Poor Underserved* 2008; **19**: 41–55.
20 Birman-Deych E, Radford MJ, Nilasena DS, Gage BF. Use and effectiveness of warfarin in Medicare beneficiaries with atrial fibrillation. *Stroke* 2006; **37**: 1070–4.
21 Shen AY, Yao JF, Brar SS, Jorgensen MB, Wang X, Chen W. Racial/ethnic differences in ischemic stroke rates and the efficacy of warfarin among patients with atrial fibrillation. *Stroke* 2008; **39**: 2736–43.
22 Fang MC, Machtinger EL, Wang F, Schillinger D. Health literacy and anticoagulation-related outcomes among patients taking warfarin. *J Gen Intern Med* 2006; **21**: 841–6.
23 Palareti G, Legnani C, Guazzaloca G, Lelia V, Cosmi B, Lunghi B, Marchetti G, Poli D, Pengo V. Risk factors for highly unstable response to oral anticoagulation: a case-control study. *Br J Haematol* 2005; **129**: 72–8.
24 Iezzoni LI. *Risk Adjustment for Measuring Health Care Outcomes*, 3rd edn. Chicago, IL: Health Administration Press, 2003.
25 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**: 857–67.
26 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**: 1139–51.
27 Schulman S, Kearon C, Kakkar AK, Mismetti P, Schellong S, Eriksson H, Baanstra D, Schnee J, Goldhaber SZ. Dabigatran versus warfarin in the treatment of acute venous thromboembolism. *N Engl J Med* 2009; **361**: 2342–52.
28 O'Brien CL, Gage BF. Costs and effectiveness of ximelagatran for stroke prophylaxis in chronic atrial fibrillation. *JAMA* 2005; **293**: 699–706.
29 Wasson JH, Sox HC, Neff RK, Goldman L. Clinical prediction rules. Applications and methodological standards. *N Engl J Med* 1985; **313**: 793–9.
30 Laupacis A, Sekar N, Stiell IG. Clinical prediction rules. A review and suggested modifications of methodological standards. *JAMA* 1997; **277**: 488–94.

Anticoagulation for valvular heart disease in community-based practice

Adam J. Rose[1,2]; Al Ozonoff[1,3]; Lori E. Henault[2]; Elaine M. Hylek[2]

[1]Center for Health Quality, Outcomes and Economic Research, Bedford VA Medical Center, Bedford, Massachusetts, USA; [2]Department of Medicine, Section of General Internal Medicine, Boston University School of Medicine, Boston, Massachusetts, USA; [3]Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, USA

Summary

Little is known about patients who receive oral anticoagulation for valvular heart disease (VHD) in community-based practice. It was this study's objective to describe the characteristics, management, and outcomes of patients anticoagulated for VHD, compared to patients anticoagulated for atrial fibrillation (AF). We used a nationally-representative cohort of community-based anticoagulation care in the United States. Data collected included indications for therapy, demographics, selected comorbid conditions, international normalised ratio (INR) target ranges, INR control, and clinical outcomes. We identified 1,057 patients anticoagulated for VHD (15.6% of the overall cohort) and 3,396 patients anticoagulated for AF (50.2%). INR variability was similar between the two groups (0.64 vs. 0.69, p = 0.80). Among patients with aortic VHD, for whom a standard (2–3) target INR range is recommended, 461 (84%) had a high target range (2.5–3.5), while 95 (16%) had a standard target range. VHD patients had a higher rate of major haemorrhage compared to AF patients (3.57 vs. 1.78 events per 100 patient-years, incidence rate ratio 2.02, 95% CI 1.33 – 3.06). The rate of stroke/systemic embolus was similar between groups (0.67 vs. 0.97 events per 100 patient-years, incidence rate ratio 0.71, 95% CI 0.32 – 1.57). In our community-based study, approximately 15.6% of patients receiving warfarin were anticoagulated for VHD. VHD patients achieved similar anticoagulation control to patients with AF, as measured by INR variability. Nevertheless, the rate of major haemorrhage was elevated among VHD patients compared to AF patients; this finding requires further investigation.

Keywords

Warfarin, anticoagulants, quality of health care, heart valve diseases

Correspondence to:
Adam J. Rose, MD MSc
Center for Health Quality, Outcomes, and Economic Research
Bedford VA Medical Center
200 Springs Road, Building 70
Bedford, MA 01730, USA
Tel.: +1 781 687 3198, Fax: +1 917 591 3104
E-mail: adamrose@bu.edu

Disclaimer: The opinions expressed in this manuscript do not necessarily represent the views or policies of the United States Department of Veterans Affairs.

Financial support:
This study was funded by Bristol-Myers Squibb, the makers of Coumadin® brand warfarin. Bristol-Myers Squibb had no role in the design and conduct of the study, or in the collection, analysis, and interpretation of the data, or in the preparation, review, and approval of the manuscript. Dr. Rose is supported by a career development award from the United States Department of Veterans Affairs Health Services Research and Development Service.

Received: July 14, 2009
Accepted after major revision: October 23, 2009
Prepublished online: December 18, 2009

doi:10.1160/TH09-07-0450
Thromb Haemost 2010; 103: 329–337

Introduction

The three major indications for long-term anticoagulation with warfarin are atrial fibrillation (AF) (1–2), venous thromboembolism (3), and valvular heart disease (VHD)(4). Of these three categories, we know the least about patients who are anticoagulated for VHD, a heterogeneous category that includes bioprosthetic valve replacements, mechanical valve replacements, valve repairs, and various other kinds of VHD. There is evidence to support the efficacy of anticoagulation in preventing strokes in this population, and considerable information about the expected rate of thromboembolic events with and without anticoagulation (5–23). For example, among patients with heart valve prosthesis not receiving anticoagulation, the yearly rate of thromboembolic events is approximately 12% with an aortic prosthesis and 22% with a mitral prosthesis (23). Various studies have found that anticoagulation lowers this rate to approximately 1–2% per year (4). Given this large effect size, prosthetic heart valves and other VHD conveying a similar risk of thromboembolism are considered a near-absolute indication for long-term anticoagulation.

However, previous studies have not elucidated the characteristics of patients who receive anticoagulation for VHD in real-life clinical practice, including their demographics, comorbid conditions, precise indications for therapy, target international normalised ratio (INR) ranges, anticoagulation control, and outcomes. One large cohort, the ATRIA study, only included patients with AF (24). Another cohort, the ISCOAT study (25), enrolled 479 patients with VHD (17.5% of the cohort), but did not analyse them separately. Other important cohorts (26–27) did not explicitly report the proportion of patients anticoagulated for VHD.

Our objective, therefore, was to characterise the care of patients anticoagulated for VHD in community-based practice using a large, nationally-representative database of anticoagulation care in the United States. This study can provide important information regarding the proportion of patients who receive anticoagulation for VHD, the relative proportion of sub-categories of VHD, the demographics and comorbid illnesses of VHD patients, the target INR ranges chosen for patients with VHD, the anticoagulation control achieved by patients with VHD, use of low-molecular-weight heparin (LMWH) to "bridge" periods of low INR, and rates of thromboembolic and haemorrhagic events in routine clinical practice.

Materials and methods

Study enrolment

Data collection for the Anticoagulation Consortium to Improve Outcomes Nationally (ACTION) study has been described in detail elsewhere (28). The objective of ACTION was to collect a representative cohort of anticoagulation patients managed in community-based settings, in order to facilitate outcomes research and health services research on important anticoagulation-related topics. Physician practices that were registered users of CoumaCare® software (Bristol-Myers Squibb, Princeton, NJ, USA) were invited to participate. In total, 174 practices registered online to participate and 101 sites had the technologic capability and the review board approval necessary to proceed. The 101 study sites were from diverse geographic regions, representing 34 states in the United States. All sites had at least one dedicated provider managing warfarin, usually within the setting of a community-based, physician group practice.

Patients were invited to participate by letter, clinic flyer, or in person (at the time of a routine appointment). To be eligible, patients had to be 18 years of age or older and provide written informed consent. Enrolment began in April 2000 and follow-up ended in March 2002. Encrypted data from the electronic patient anticoagulation record were downloaded to the data-coordinating centre weekly. Missing data fields and data entry errors were flagged and resolved directly with the sites by the data-coordinating centre before data were transferred to the study investigators. The study was approved by Western Institutional Review Board® (WIRB®) of Olympia, WA, USA and by local review boards where they existed.

Study variables

Data collected included patient demographics, the presence or absence of several important comorbid conditions (congestive heart failure, coronary artery disease, diabetes, hypertension, and prior stroke), INR target ranges, INR values, and patient management notes. We characterised INR control using the INR variability measure described by Fihn et al. (26, 29) This measure computes the mean INR value for each patient, and then uses the standard deviation (SD) around this mean as a measure of INR variability over time. We chose to use INR variability in this study because patients with VHD may have either standard (2 – 3) or high (2.5 – 3.5) INR target ranges, and the calculation of INR variability is independent of the INR target range. However, we also calculated percent time in therapeutic range (TTR), as originally described by Rosendaal (30). For this calculation of TTR, patients were assigned a target INR range of 2.0 – 3.0 ("standard") or 2.5 – 3.5 ("high") according to the target range indicated in the clinical record.

Indications for anticoagulation

Patients in the database had a recorded primary indication for therapy and in some cases a secondary indication as well. These were confirmed by review of the clinical notes. When a patient had VHD and another indication for therapy (e.g. AF), VHD was considered the primary indication. VHD patients were further subdivided by the anatomic location and nature of the valvular abnormality. When patients had both aortic and mitral valve disease, the mitral valve abnormality was considered the primary indication for therapy. In some cases, the anatomic location of the VHD (i.e. aortic vs. mitral position) could not be determined. We also compared these patients with VHD to the 3,396 patients in the ACTION study anticoagulated for AF, who have been described in a previous report (28). Patients with AF form a useful comparison group, since the management and outcomes of anticoagulation for AF has been studied extensively and is comparatively well-understood.

Low INR values and bridging with LMWH

We identified all INR values 1.5 or lower among patients anticoagulated for VHD, and used linear interpolation (30) to identify periods of time when the INR was ≤1.5. For each low INR value, we reviewed the clinical notes to determine whether clinicians employed "bridging" with heparin or LMWH as part of the management of the low INR value, and to characterise the circumstances surrounding such episodes.

Adverse events

Ischaemic stroke, systemic arterial embolism, and major haemorrhage were the adverse outcomes of interest. All events were confirmed by manual review of the clinical notes in the database. Major haemorrhage was defined according to the International Society on Thrombosis and Haemostasis (ISTH) definition (31): a

fatal event, an event requiring hospitalisation 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. All patient progress notes were individually reviewed for evidence of adverse events; events were validated directly with the sites by the data coordinating centre.

Statistical analyses

We compared the VHD and AF groups with regard to demographics, comorbid conditions, frequency of INR testing, frequency of intentional interruptions of warfarin therapy, INR variability, and percent time in the therapeutic INR range (TTR). Within the VHD group, we investigated patient-level factors associated with assignment to the standard or the high target INR range, using bivariate analyses. To account for intra-class correlation within clinical site, p-values for continuous variables were calculated using Generalized Estimating Equations (GEE), while p-values for nominal variables were calculated using 10,000 Monte Carlo simulations. We compared rates of bridging per 100 low INR values between the VHD and AF groups. We computed the rate of stroke/systemic embolus among VHD patients and estimated the proportion of these events attributable to periods of low INR. We computed the rate of major haemorrhage among VHD patients, and estimated the proportion of these events attributable to the use of bridging. When comparing rates of bridging or clinical events between the VHD and AF groups, we computed incidence rate ratios with 95% confidence intervals (CI) via bootstrapping with 10,000 iterations.

We compared the rates of these adverse events among our VHD cohort to the rates we previously described among a cohort of AF

Table 1: Comparison of patients anticoagulated for valvular heart disease (VHD) and for atrial fibrillation (AF).

	Valvular heart disease (n = 1057)	Atrial fibrillation (n = 3396)	P-value
Mean age (SD)	67.2 (12.2)	74.1 (9.4)	< 0.001
Female gender	460 (44%)	1433 (42%)	0.44
Race			
White	959 (91%)	3182 (94%)	0.04
Black	23 (2%)	47 (1%)	
Other	75 (7%)	167 (5%)	
Comorbid conditions			
Congestive heart failure	216 (20%)	642 (19%)	0.37
Coronary artery disease	247 (23%)	855 (25%)	0.80
Diabetes	139 (13%)	508 (15%)	0.14
Hypertension	395 (37%)	1530 (45%)	< 0.001
Prior stroke	104 (10.0%)	359 (11%)	0.52
Measures of warfarin management			
Mean days in database (SD)	358 (124)	328 (133)	< 0.001
Mean number of INR values (SD)	21 (11)	16 (8)	< 0.001
Mean INR values/month (SD)	1.79 (0.84)	1.65 (0.90)	< 0.001
Intentional interruptions of warfarin for procedures per patient-year (95% CI)	0.35 (0.29, 0.39)	0.43 (0.35, 0.44)	0.001
Anticoagulation control – INR variability			
Mean INR value	2.93	2.45	< 0.001
Variability around mean value	0.64	0.69	0.80
Anticoagulation control – time in range*			
Percent time below target range	21.5 (18.9)	20.1 (18.8)	0.02
Percent time in target range (TTR)	64.3 (19.6)	66.6 (19.7)	0.003
Percent time above target range	14.5 (15.5)	13.3 (13.5)	0.22

To account for intra-class correlation within clinical site, p-values for continuous variables were calculated using Generalized Estimating Equations (GEE), while p-values for nominal variables were calculated using 10,000 Monte Carlo simulations. *Time in Range is calculated using a target INR range of 2.0 – 3.0 for patients with a standard target range and 2.5 – 3.5 for patients with a high target range. Some patients with valvular heart disease had a high target INR range, but no patients with atrial fibrillation had a high target range.

patients drawn from the same database (28), computing separate rates as well for periods when the INR was above, below, or within the target range. Within the VHD group, time within the target range was calculated according to whether the patient had a high (2.5 – 3.5) or normal (2 – 3) target range. Calculation of incidence rates and their confidence intervals were accomplished by Poisson regression for count data, run via GEEs in order to account for the correlation of data by site of care. Analyses were performed using the R statistical package, version 2.8 (R Foundation, 2009).

Results

Study cohort

There were 1,116 patients anticoagulated for VHD in the ACTION database. Of these, 28 patients had bioprosthetic heart valves; such patients ordinarily require only a brief period of anticoagulation following surgery, unless there is another risk factor for stroke such as AF (4). Because these patients were so few, and were not representative of most patients with VHD, they were excluded from our study. In addition, we excluded 31 patients with two or fewer INR values in the database, too few to characterise their care. After these exclusions, our final VHD cohort included 1,057 patients (15.6% of the entire cohort of 6,761 patients, and 94.7% of the original 1,116 VHD patients). By comparison, there were 3,396 patients anticoagulated for AF in the ACTION cohort (50.2%).

Among VHD patients, 556 (53%) had a disorder of the aortic valve, while 417 (39%) had a disorder of the mitral valve (with or without a concomitant aortic valve disorder). For 84 (8%), the anatomic location of the valvular heart condition was unclear from the clinical record. As to the valve pathology necessitating the anticoagulation, 952 (90%) had heart valve prostheses, while 105 (10%) had other VHD.

VHD patients compared to AF patients

VHD patients were compared to AF patients, a group which has been well-characterised in previous studies (▶Table 1). VHD patients were considerably younger than AF patients (mean age 67.2 vs. 74.1, $p < 0.001$). The prevalence of several important comorbid conditions (congestive heart failure, coronary artery disease, diabetes, and prior stroke) was similar between the two groups. However, hypertension was somewhat more common in the AF group (45% vs. 37%, $p < 0.001$).

Patients with VHD had their INR tested somewhat more often (1.79 vs. 1.65 tests/month, $p < 0.001$), and had fewer intentional interruptions of warfarin therapy for procedures per patient-year (0.35 vs. 0.43, $p = 0.001$). The mean INR value was higher in the VHD group than the AF group (2.93 vs. 2.45, $p < 0.001$), reflecting the many VHD patients with a high target range. Patients with AF achieved a slightly higher percentage of time in therapeutic range (66.6% vs. 64.3%), a difference which was statistically significant due to the large sample size. However, the two groups had similar INR variability (0.64 vs. 0.69, $p = 0.80$), suggesting that the two groups experienced similar INR control.

Standard vs. high target INR range among VHD patients

Most VHD patients (84%) had a high target INR range (i.e. 2.5 – 3.5), as opposed to a standard target range (i.e. 2.0 – 3.0). We compared VHD patients with a standard target range to those with a high target range (▶Table 2). The mean INR values achieved in the two groups (2.51 vs. 3.00, $p < 0.001$) suggest that the target ranges stated in the record were actually pursued in practice. Patients with a standard target range were older, a difference of marginal statistical significance (mean age 69.6 vs. 66.7, $p = 0.053$), but the two groups were similar in terms of gender, race, and comorbid conditions. Patients had a similar likelihood of assignment to a high target range regardless of whether their VHD was in the aortic or the mitral anatomic position (83% vs. 86%, $p = 0.28$).

Low INR values and bridging in patients with VHD

Bridging was more commonly employed among patients anticoagulated for VHD than for AF, although most episodes of low INR did not trigger bridging even in the VHD group. VHD patients had 8.1 episodes of bridging per 100 low INR values (95% CI 6.0 – 10.5), while AF patients had a rate of 1.1 (95% CI 0.8 – 1.5, $p < 0.001$ for a between-group difference). Of the 39 episodes of bridging that occurred among the VHD cohort, the great majority (34 episodes) were in response to intentional interruptions of therapy for a procedure ("holds"), while only five were in response to routinely encountered (unplanned) low INR values.

Sites differed widely in their practice regarding bridging. For example, with regard to VHD patients, of 79 sites that had at least one episode of low INR, there were 60 sites (76%) with no episodes of bridging. The rate among the 19 sites that did employ bridging was 15.0 episodes per 100 low INR values (95% CI 11.2 – 19.5).

Rates of adverse events

In the VHD cohort, there were seven ischaemic strokes (and no systemic emboli) in 1038.2 patient-years of follow-up, an incidence rate of 0.67 events per 100 patient-years (95% CI 0.32 – 1.43). The overall rate of stroke/systemic emboli in the VHD cohort was similar to the rate of such events among AF patients in this database (0.97 events per 100 patient-years). The incidence rate ratio between the two groups (comparing VHD to AF) was 0.71 (95% CI 0.32 – 1.57).

Among the VHD cohort, there were 37 major haemorrhagic events in 1038.2 patient-years of follow up (3.57 events per 100 patient-years, 95% CI 2.59 – 4.92). Three of these events (an intraperitoneal haemorrhage, a gastrointestinal haemorrhage, and a retroperitoneal haemorrhage) occurred during episodes of bridging. Among our AF cohort, the rate of major haemorrhage was 1.78 events per 100 patient-years. The incidence rate ratio (comparing VHD to AF) was 2.02 (95% CI 1.33 – 3.06), a difference that was statistically significant at the 0.05 level.

Rates of adverse events, stratified by indication for therapy and INR value at the time of the event, are given in ▶Table 3. The rate of major haemorrhage was higher among VHD patients than among AF patients only while the INR was below or within the target range, but not above. However, this finding should be considered in light of the fact that most patients in the VHD group had a high target range (2.5 – 3.5). Of the 13 major haemorrhagic events that occurred in the VHD cohort during time below the target range, 10 occurred while the INR was between 2 – 2.5. Similarly, of

Table 2: Correlates of high vs. standard target INR range in patients anticoagulated for valvular heart disease (VHD). Also, a comparison of anticoagulation control between the two groups.

	Standard target (2 – 3) (n = 174)	High target (2.5 – 3.5) (n = 883)	P-value
Group mean INR value (SD)	2.51 (0.27)	3.00 (0.36)	< 0.001
Mean age (SD)	69.6 (12.2)	66.7 (12.1)	0.053
Female gender	69 (40%)	391 (44%)	0.28
Race			**0.43**
White	156 (90%)	803 (91%)	
Black	2 (1%)	21 (2%)	
Other	16 (9%)	59 (7%)	
Comorbid conditions			
Congestive heart failure	34 (20%)	182 (21%)	0.88
Coronary artery disease	49 (28%)	198 (22%)	0.36
Diabetes	30 (17%)	109 (12%)	0.11
Hypertension	77 (44%)	318 (36%)	0.38
Prior stroke	11 (6%)	93 (10%)	0.09
Anatomic valve location			**0.02**
Mitral valve (with or without another)	58 (32%)	359 (40%)	
Aortic valve	95 (52%)	461 (51%)	
Unknown	21 (11%)	63 (7%)	
Type of valve			**< 0.001**
Prosthetic heart valve	111 (64%)	841 (95%)	
Other valve disorder or unknown	63 (36%)	42 (5%)	
Secondary indications for anticoagulation			0.02
Atrial fibrillation	60 (34%)	206 (23%)	
Prior stroke or systemic embolus	2 (1%)	47 (5%)	
None	112 (64%)	630 (71%)	
Anticoagulation control – INR variability			
Group Mean INR Value	2.51 (0.27)	3.00 (0.36)	< 0.001
INR variability	0.48 (0.37)	0.67 (0.69)	<0.001
Anticoagulation control – time in range*			
Percent time below target range	15.9 (18.4)	22.6 (18.8)	<0.001
Percent time in target range (TTR)	71.6 (19.4)	62.8 (19.3)	<0.001
Percent time above target range	12.5 (13.7)	14.9 (15.8)	0.03

To account for intra-class correlation within clinical site, p-values for continuous variables were calculated using Generalized Estimating Equations (GEE), while p-values for nominal variables were calculated using 10,000 Monte Carlo simulations. *Time in Range is calculated using a target INR range of 2.0 – 3.0 for patients with a standard target range and 2.5 – 3.5 for patients with a high target range.

	Valvular heart disease (n = 1057)	Atrial fibrillation (n = 3396)	P-value
Time below target range			
Patient-years	166.7	560.5	--
Stroke/systemic embolus events	4	11	--
Incidence Rate (95% CI)	2.91 (1.19, 7.12)	1.96 (1.07, 3.60)	0.71
Major haemorrhage events	13	16	--
Incidence rate (95% CI)	7.78 (4.49, 13.51)	2.45 (1.26, 4.76)	0.03
Time within target range			
Patient-years	537.9	1993.7	--
Stroke/systemic embolus events	2	18	--
Incidence rate (95% CI)	0.37 (0.09, 1.48)	0.87 (0.52, 1.44)	0.25
Major haemorrhage events	19	20	--
Incidence rate (95% CI)	3.76 (2.58, 5.47)	0.95 (0.58, 1.55)	<0.001
Time above target range			
Patient-years	115.7	379.4	--
Stroke/systemic embolus events	1	0	--
Incidence rate (95% CI)	0.88 (0.13, 6.05)	0.00	NA
Major haemorrhage events	5	18	--
Incidence rate (95% CI)	3.80 (1.42, 10.18)	4.66 (2.83, 7.66)	0.78

Incidence rates are expressed as events per 100 patient-years. *Time in range is calculated using a target INR range of 2.0 – 3.0 for patients with a standard target range and 2.5 – 3.5 for patients with a high target range. Some patients with valvular heart disease had a high target INR range, but no patients with atrial fibrillation had a high target range.

Table 3: Adverse events and relationship to anticoagulation control among patients anticoagulated for valvular heart disease (VHD) and atrial fibrillation (AF).

the 19 major haemorrhagic events that occurred in the VHD cohort during time within the target range, seven occurred while the INR was between 3 – 3.5.

Discussion

To our knowledge, this is the first study to describe the patient population receiving long-term oral anticoagulation for VHD in community-based practice. VHD patients composed 15.6% of our community-based anticoagulation cohort, a similar proportion to that seen in a previous Italian anticoagulation cohort (25). We describe the different types of VHD represented in our VHD cohort. Most patients (90%) were anticoagulated because of a prosthetic valve, while only 10% of patients had other VHD. Compared to patients with AF, VHD patients were considerably younger, but otherwise similar in terms of demographics and most comorbid conditions. This younger age among VHD patients is probably due to the fact that older patients requiring valve replacement would have received bioprosthetic valves, which require only a brief period of anticoagulation after surgery (4). Despite the fact that most (84%) VHD patients had a high target INR range, their INR variability was similar to that of the AF group, suggesting that their INR was similarly well-controlled.

Despite similar anticoagulation control in both groups, VHD patients experienced a higher rate of major haemorrhage than AF patients. Stratification of these events into those that occurred within different INR ranges (▶Table 3) suggests that the use of a high target INR range for many of the VHD patients may have played a role in this between-group difference. Another possible reason for this finding is that clinical guidelines, both now (4) and at the time of the study (32–33), recommend adding low-dose aspirin to warfarin for many patients with prosthetic heart valves. This recommendation was based on several well-conducted trials and meta-analyses, which suggested that the addition of low-dose aspirin for patients already receiving warfarin could improve outcomes like mortality with only a minimal increase in the rate of major haemorrhage (34–36). The guidelines caution, however, that patients at elevated risk for haemorrhage probably should not receive aspirin in addition to warfarin (4, 32–33), because the risk may outweigh the benefit. We searched the anticoagulation record of all VHD patients that experienced major haemorrhage for a mention of aspirin therapy. Four patients were clearly receiving aspirin at the time of a major haemorrhage, and a fifth patient was receiving clopridogrel at the time of her two major haemorrhagic events. The other charts contained no evidence of concomitant antiplatelet therapy. This is a limitation of our study, because the absence of documentation is not conclusive.

Regarding target INR ranges, the overwhelming majority of VHD patients in our sample (84%) had a high INR target range, whether their valve pathology was in the mitral or the aortic position. This is surprising, because clinical guidelines both now (4) and at the time of the study (33, 37–39) would have recommended a standard target range for patients with aortic prosthetic valves, except for the three subgroups: those with concomitant AF, those with prior stroke despite adequate anticoagulation, and those with caged-ball prosthetic valves. Caged-ball valves, which were first implanted in the 1960s, were quickly supplanted after the introduction of newer, less thrombogenic valves in the early 1970s (40). Since our study began in 2000, it is doubtful that many of our patients had these older caged-ball valves.

However, it is possible that the other two subgroups accounted for some of our findings. Therefore, we identified all of the patients with aortic valve prostheses who had prior stroke/systemic embolus and/or AF. For the purpose of this sensitivity analysis, we assumed that all prior strokes had occurred while the patient was receiving warfarin and was within the therapeutic range. Of the 556 patients with aortic valve prostheses, 130 had one or both of these possible reasons for a high target range. Of these 130 high-risk patients, 82% had a high target range, compared to 84% of the patients without these additional risk factors ($p = 0.60$ for comparison). This non-differential management suggests that clinicians in our study were unaware of or did not agree with clinical guideline recommendations regarding the recommended target INR range for patients with a prosthetic aortic valve (41). As discussed above, it is possible that this guideline-discordant use of a high target range contributed to the higher rate of major haemorrhage observed in the VHD group in our study. The selection of target INR ranges for VHD patients might be improved through education and outreach.

Regarding the management of low INR values, our study suggests that "bridging" is employed much more often for VHD patients than AF patients, almost always in response to a planned interruption of therapy. In a previous analysis of the ACTION database, we characterised the use of bridging for intentional interruptions of therapy, as well as the risks of haemorrhage and thromboembolism with and without bridging (42). However, this previous study did not examine patients with VHD, who may have the highest risk of thromboembolism with therapy interruptions, and thus the strongest rationale for bridging. Another study, performed using a different database, examined the risk of thromboembolism during the 90 days after a low INR value among patients with diverse indications for anticoagulation, although there were few VHD patients in that study (4%). That study suggested that the risk of thromboembolism attributable to a low INR value, compared to patients who did not have a low INR, is minimal (43).

In this study, we observed considerable variations in practice regarding bridging. Several sites in our sample used bridging relatively often for VHD patients, while most sites (78%) did not use it at all. This variation in practice reflects a lack of evidence and consensus regarding this difficult clinical decision (44), which requires clinicians to balance considerable risks and benefits in the absence of empiric evidence (1–2, 4, 45). An ongoing randomised trial of bridging is expected to help clarify the risks and benefits of bridging (46). However, this trial will only enroll patients anticoagulated for AF who have intentional interruptions of therapy, so it will not directly inform decisions for patients with VHD or patients with incidentally noted low INR values.

This study has several important strengths. Our database was large and nationally-representative, drawn from 101 clinical sites in 38 states. It therefore provides a generalisable picture of VHD patients in community-based practice. Our database also contained a degree of clinical detail that would not be possible with automated data alone.

However, this study also had some limitations. First, most of the patients in the ACTION study were prevalent users of warfarin as opposed to an inception cohort. It is well-known that patients new to warfarin have poorer anticoagulation control and higher rates of complications than experienced users (28, 47). Therefore, the relatively good anticoagulation control and low rates of adverse events reported in this study may not be generalisable to patients new to warfarin. Second, we computed percent time in therapeutic range (TTR) for patients with different target INR ranges (2.0 – 3.0 and 2.5 – 3.5), although TTR has not explicitly been demonstrated to mean the same thing in patients with different target ranges. However, the group that invented the TTR measure has used it to measure anticoagulation control in patients with varying target INR ranges (30, 48), so this assumption may be justified. Third, while our results suggest that many patients with aortic VHD were inappropriately treated with a high target range, we cannot fully exclude the possibility that some patients had a valid reason for a high target range which was not recorded in the clinical notes.

What is known about this topic?

- Warfarin is indicated for the prevention of stroke in patients with valvular heart disease or prosthetic heart valves (VHD).
- The risk of stroke is higher with VHD than with atrial fibrillation (AF). Therefore, VHD is considered an almost absolute indication for anticoagulation, while the decision to anticoagulate patients with AF is based on a risk-benefit analysis.
- Previous studies of anticoagulation for VHD have been clinical trials conducted in academic medical centers. There have been no large cohort studies describing patients anticoagulated for VHD in community-based practice.

What does this paper add?

- This paper describes the care and outcomes of 1,057 patients anticoagulated for VHD at 101 community-based sites of care located in 31 states of the United States. Approximately 15.6% of patients receiving warfarin at these sites of care were anticoagulated for VHD.
- Despite similar levels of anticoagulation control between the two groups, patients anticoagulated for VHD had a higher rate of major haemorrhage than patients anticoagulated for AF (incidence rate ratio 2.02, 95% CI 1.33 – 3.06). Further studies are needed to investigate the cause of this between-group difference in outcomes.

Similarly, there were some patients in our study whose VHD was not completely characterised regarding anatomic position and/or type of prosthetic valve. Furthermore, our database lacked explicit detail regarding the causes of valvular heart disease (rheumatic, etc.), the exact type of prosthesis (St. Jude, Bjork-Shiley, etc.), the exact date of cardiac surgery, and the exact date of initiation of anticoagulation. Finally, as noted earlier, we found few references to concomitant antiplatelet therapy in the charts of VHD patients with major haemorrhagic events, but the absence of documentation is not conclusive. Complete data regarding antiplatelet therapy might have allowed further elucidation of the elevated risk of major haemorrhage that we observed for VHD patients.

In conclusion, 15.6% of patients taking warfarin in community-based practice receive anticoagulation for VHD. VHD patients are considerably younger than AF patients, but the two groups are otherwise similar in many respects, including comorbid conditions and degree of anticoagulation control. Community practice regarding LMWH for patients with VHD for episodes of low INR ("bridging") is highly variable, reflecting a lack of evidence and clinical consensus. Finally, patients with VHD experienced a greatly elevated risk of major haemorrhage in this study, compared to patients with AF. Further study will be necessary to confirm and explain this finding.

References

1. Fuster V, Ryden LE, Cannom DS, et al. ACC/AHA/ESC 2006 Guidelines for the Management of Patients with Atrial Fibrillation. Circulation 2006; 114: e257–354.
2. 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: 546S-592S.
3. Kearon C, Kahn SR, Agnelli G, et al. Antithrombotic therapy for venous thromboembolic disease: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition). Chest 2008; 133: 454S-545S.
4. Salem DN, O'Gara PT, Madias C, et al. Valvular and structural heart disease: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition). Chest 2008; 133: 593S-629S.
5. 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: 316–320.
6. Acar J, Iung B, Boissel JP, et al. AREVA: multicenter randomized comparison of low-dose versus standard-dose anticoagulation in patients with mechanical prosthetic heart valves. Circulation 1996; 94: 2107–2112.
7. Baudet EM, Oca CC, Roques XF, et al. A 5 1/2 year experience with the St. Jude Medical cardiac valve prosthesis. Early and late results of 737 valve replacements in 671 patients. J Thorac Cardiovasc Surg 1985; 90: 137–144.
8. Bloomfield P, Wheatley DJ, Prescott RJ, et al. Twelve-year comparison of a Bjork-Shiley mechanical heart valve with porcine bioprostheses. N Engl J Med 1991; 324: 573–579.
9. Butchart EG, Payne N, Li HH, et al. Better anticoagulation control improves survival after valve replacement. J Thorac Cardiovasc Surg 2002; 123: 715–723.
10. Emery RW, Arom KV, Nicoloff DM. Utilization of the St. Jude Medical prosthesis in the aortic position. Semin Thorac Cardiovasc Surg 1996; 8: 231–236.
11. Fiore AC, Barner HB, Swartz MT, et al. Mitral valve replacement: randomized trial of St. Jude and Medtronic Hall prostheses. Ann Thorac Surg 1998; 66: 707–713.
12. Hering D, Piper C, Bergemann R, et al. Thromboembolic and bleeding complications following St. Jude Medical valve replacement: results of the German Experience With Low-Intensity Anticoagulation Study. Chest 2005; 127: 53–59.
13. Horstkotte D, Schulte H, Bircks W, et al. Unexpected findings concerning thromboembolic complications and anticoagulation after complete 10 year follow up of patients with St. Jude Medical prostheses. J Heart Valve Dis 1993; 2: 291–301.
14. Horstkotte D, Schulte HD, Bircks W, et al. Lower intensity anticoagulation therapy results in lower complication rates with the St. Jude Medical prosthesis. J Thorac Cardiovasc Surg 1994; 107: 1136–1145.
15. Laffort P, Roudaut R, Roques X, et al. Early and long-term (one-year) effects of the association of aspirin and oral anticoagulant on thrombi and morbidity after replacement of the mitral valve with the St. Jude medical prosthesis: a clinical and transesophageal echocardiographic study. J Am Coll Cardiol 2000; 35: 739–746.
16. Sethia B, Turner MA, Lewis S, et al. Fourteen years' experience with the Bjork-Shiley tilting disc prosthesis. J Thorac Cardiovasc Surg 1986; 91: 350–361.
17. Talwar S, Kapoor CK, Velayoudam D, et al. Anticoagulation protocol and early prosthetic valve thrombosis. Indian Heart J 2004; 56: 225–228.
18. Vallejo JL, Gonzalez-Santos JM, Albertos J, et al. Eight years' experience with the Medtronic-Hall valve prosthesis. Ann Thorac Surg 1990; 50: 429–436.
19. Vogt S, Hoffmann A, Roth J, et al. Heart valve replacement with the Bjork-Shiley and St Jude Medical prostheses: a randomized comparison in 178 patients. Eur Heart J 1990; 11: 583–591.
20. Koertke H, Zittermann A, Tenderich G, et al. Low-dose oral anticoagulation in patients with mechanical heart valve prostheses: final report from the early self-management anticoagulation trial II. Eur Heart J 2007; 28: 2479–2484.
21. Koertke H, Zittermann A, Wagner O, et al. Self-management of oral anticoagulation therapy improves long-term survival in patients with mechanical heart valve replacement. Ann Thorac Surg 2007; 83: 24–29.
22. Cannegieter SC, Rosendaal FR, Wintzen AR, et al. Optimal oral anticoagulant therapy in patients with mechanical heart valves. N Engl J Med 1995; 333: 11–17.
23. Baudet EM, Puel V, McBride JT, et al. Long-term results of valve replacement with the St. Jude Medical prosthesis. J Thorac Cardiovasc Surg 1995; 109: 858–870.
24. 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? J Am Med Assoc 2003; 290: 2685–2692.
25. Palareti G, Leali N, Coccheri S, et al. Bleeding complications of oral anticoagulant treatment: an inception-cohort, prospective collaborative study (ISCOAT). Italian Study on Complications of Oral Anticoagulant Therapy. Lancet 1996; 348: 423–428.
26. Fihn SD, McDonell M, Martin D, et al. Risk factors for complications of chronic anticoagulation. A multicenter study. Warfarin Optimized Outpatient Follow-up Study Group. Ann Intern Med 1993; 118: 511–520.
27. van der Meer FJ, Rosendaal FR, Vandenbroucke JP, et al. Bleeding complications in oral anticoagulant therapy. An analysis of risk factors. Arch Intern Med 1993; 153: 1557–1562.
28. Rose AJ, Ozonoff A, Henault LE, et al. Warfarin for atrial fibrillation in community-based practise. J Thromb Haemost 2008; 6: 1647–1654.
29. Kent DL, Vermes D, McDonell M, et al. A model for planning optimal follow-up for outpatients on warfarin anticoagulation. Warfarin Optimal Outpatient Follow-up Study Group. Med Decis Making 1992; 12: 132–141.
30. Rosendaal FR, Cannegieter SC, van der Meer FJ, et al. A method to determine the optimal intensity of oral anticoagulant therapy. Thromb Haemost 1993; 69: 236–239.
31. 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.
32. Stein PD, Alpert JS, Bussey HI, et al. Antithrombotic therapy in patients with mechanical and biological prosthetic heart valves. Chest 2001; 119: 220S-227S.
33. Stein PD, Alpert JS, Dalen JE, et al. Antithrombotic therapy in patients with mechanical and biological prosthetic heart valves. Chest 1998; 114: 602S-610S.
34. Cappelleri JC, Fiore LD, Brophy MT, et al. Efficacy and safety of combined anticoagulant and antiplatelet therapy versus anticoagulant monotherapy after mechanical heart-valve replacement: a metaanalysis. Am Heart J 1995; 130: 547–552.
35. Meschengieser SS, Fondevila CG, Frontroth J, et al. Low-intensity oral anticoagulation plus low-dose aspirin versus high-intensity oral anticoagulation alone: a randomized trial in patients with mechanical prosthetic heart valves. J Thorac Cardiovasc Surg 1997; 113: 910–916.
36. Turpie AG, Gent M, Laupacis A, et al. A comparison of aspirin with placebo in patients treated with warfarin after heart-valve replacement. N Engl J Med 1993; 329: 524–529.
37. Salem DN, Daudelin HD, Levine HJ, et al. Antithrombotic therapy in valvular heart disease. Chest 2001; 119: 207S-219S.
38. Salem DN, Levine HJ, Pauker SG, et al. Antithrombotic therapy in valvular heart disease. Chest 1998; 114: 590S-601S.

39. Salem DN, Stein PD, Al-Ahmad A, et al. Antithrombotic therapy in valvular heart disease--native and prosthetic: the Seventh ACCP Conference on Antithrombotic and Thrombolytic Therapy. Chest 2004; 126: 457S-482S.
40. Wikipedia. Artificial Heart Valve. Available at: http://en.wikipedia.org/wiki/artificial_heart_valve. Last accessed: October 15, 2009.
41. Cabana MD, Rand CS, Powe NR, et al. Why don't physicians follow clinical practice guidelines? A framework for improvement. J Am Med Assoc 1999; 282: 1458–1465.
42. Garcia DA, Regan S, Henault LE, et al. Risk of thromboembolism with short-term interruption of warfarin therapy. Arch Intern Med 2008; 168: 63–69.
43. Clark NP, Witt DM, Delate T, et al. Thromboembolic consequences of subtherapeutic anticoagulation in patients stabilized on warfarin therapy: the low INR study. Pharmacotherapy 2008; 28: 960–967.
44. Garcia DA, Ageno W, Libby EN, et al. Perioperative anticoagulation for patients with mechanical heart valves: a survey of current practice. J Thromb Thrombolysis 2004; 18: 199–203.
45. 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: 160S-198S.
46. National Institutes of Health: Effectiveness of Bridging Anticoagulation for Surgery (The BRIDGE Study). ClinicalTrials.gov identifier: NCT00786474. Available at: www.clinicaltrials.gov. Last accessed: October 15, 2009.
47. 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.
48. 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.

ARTICLE | Annals of Internal Medicine

The Spread of Alcohol Consumption Behavior in a Large Social Network

J. Niels Rosenquist, MD, PhD; Joanne Murabito, MD; James H. Fowler, PhD; and Nicholas A. Christakis, MD, PhD

Background: Alcohol consumption has important health-related consequences and numerous biological and social determinants.

Objective: To explore quantitatively whether alcohol consumption behavior spreads from person to person in a large social network of friends, coworkers, siblings, spouses, and neighbors, followed for 32 years.

Design: Longitudinal network cohort study.

Setting: The Framingham Heart Study.

Participants: 12 067 persons assessed at several time points between 1971 and 2003.

Measurements: Self-reported alcohol consumption (number of drinks per week on average over the past year and number of days drinking within the past week) and social network ties, measured at each time point.

Results: Clusters of drinkers and abstainers were present in the network at all time points, and the clusters extended to 3 degrees of separation. These clusters were not only due to selective formation of social ties among drinkers but also seem to reflect interpersonal influence. Changes in the alcohol consumption behavior of a person's social network had a statistically significant effect on that person's subsequent alcohol consumption behavior. The behaviors of immediate neighbors and coworkers were not significantly associated with a person's drinking behavior, but the behavior of relatives and friends was.

Limitations: A nonclinical measure of alcohol consumption was used. Also, it is unclear whether the effects on long-term health are positive or negative, because alcohol has been shown to be both harmful and protective. Finally, not all network ties were observed.

Conclusion: Network phenomena seem to influence alcohol consumption behavior. This has implications for clinical and public health interventions and further supports group-level interventions to reduce problematic drinking.

Primary Funding Source: National Institutes of Health and the Robert Wood Johnson Foundation. The Framingham Heart Study is supported by the National Heart, Lung, and Blood Institute.

Ann Intern Med. 2010;152:426-433. www.annals.org
For author affiliations, see end of text.

Alcohol use is common in the United States. In 2002, 55% of adults reported having had at least 1 drink in the previous month, and the prevalence of past-month alcohol consumption was somewhat higher for men (62%) than for women (48%) (1). The lifetime prevalence of alcohol use disorders has been measured at 14.6% (1). Excessive alcohol use, either in the form of heavy drinking or binge drinking, increases the risk for numerous health and social problems (2, 3), and approximately 75 000 deaths in 2001 were attributable to excessive alcohol use, which makes it the third-leading lifestyle-related cause of death (3).

Alcohol consumption behavior has many determinants. Previous studies (3, 4) suggest that biological factors have a significant effect on the progression from experimentation to regular use and that social and cultural factors play a critical role in experimentation with alcohol and the development of drinking patterns over time. Given the social nature of this behavior, it is not surprising that previous work has identified interactions with friends and family members as key factors (4–8). Although this literature primarily focused on cross-sectional panels, some studies (6–8) have attempted to test whether social influences act over time. These studies, which focused on peer influence among college students, showed inconsistent results and tended to focus just on pairs of connected persons.

The study of social influences on behavior has expanded in recent years to the study of networks of linked individuals over time (9). Recent work in this area has shown that various health-related phenomena, ranging from sexually transmitted diseases to obesity, smoking, and even suicide, may travel along and within social networks (10–15).

Using a longitudinal, dynamic network of 12 067 persons, we analyzed the role of social networks in alcohol use, focusing on 1) whether clusters of heavy drinkers and abstainers existed within the network; 2) whether a person's alcohol consumption behavior was associated with that of his or her social contacts; 3) the extent to which such associations depended on the nature and direction of the social ties (for example, friends of different kinds, siblings, spouses, coworkers, or neighbors); and 4) whether gender affected the spread of alcohol consumption across social ties.

See also:

Print
Editors' Notes 427
Summary for Patients I-36

Web-Only
Appendix
Appendix Figure
Conversion of graphics into slides

METHODS

Source Data

We used data from participants in the Framingham Heart Study (FHS). The FHS is a population-based, longitudinal, observational cohort study that was initiated in 1948 to prospectively investigate risk factors for cardiovascular disease. Four cohorts, who mostly represent different generations linked to an original cohort, are included in the entire FHS. Participant data, collected every 2 to 4 years, includes physical examinations, laboratory tests, noninvasive cardiac and vascular testing, battery testing (such as the Mini-Mental State Examination), questionnaire results, and basic demographic information. For our analyses, we aligned the examination waves for the original cohort with those of the second-generation offspring cohort, which allowed us to treat all participants as having been examined in 7 waves. The offspring cohort, initiated in 1971, is the source of our study's principals, or focal individuals in the network (16). However, we included other FHS participants whom the principals listed as social contacts and refer to them here as "contacts." Therefore, even though principals come only from the offspring cohort, contacts are drawn from the entire set of both the original and offspring cohorts.

To ascertain social network ties, we created a separate data set that linked individuals through self-described social ties, collected in each of the 7 waves of the study. We could then detect relationships between participants (for example, spouse, sibling, friend, coworker, or neighbor) and observe changes in these ties across time. Either party to a link between 2 people might identify his or her link to the other. This is most relevant to the "friend" link, which could exist if A nominated B or B nominated A as a friend. We also used complete records of participants' and their contacts' address in each wave since 1971 in our analyses, although we have no information about relationships that participants did not report. For each wave, we could determine who is whose neighbor and the geographic distance between persons (10, 17). **Table 1** provides descriptive statistics for the 5124 principals in our sample.

Measures

Alcohol consumption was self-reported in all studied waves, with participants reporting their average number of drinks per week over the past year as well as the number of days within the past week during which they consumed alcohol (beer, wine, and liquor). Self-reported data are generally considered a valid and reliable source when assessing alcohol consumption, although recall measures, such as those used in this study, can be subject to recall bias from participants (18).

We treated alcohol consumption as a continuous variable in some analyses (for example, number of drinks per day, calculated from participant responses) but conducted others with dichotomous cut-points, defining heavy drinkers as those who averaged more than 1 (for women) or 2 (for men) drinks per day; moderate drinkers as those whose alcohol consumption was less than the cutoff values for heavy drinkers; and abstainers as those who reported no alcohol consumption. We did not use self-reported number of days drinking in the past week as a measure in and of itself but rather as a means to calculate average number of drinks in a day. (These labels do not reflect clinical definitions of alcohol abuse or dependence.) **Table 2** shows averages for the study population across time, including age, alcohol consumption, and percentages of abstainers and drinkers. Although the differences in how we measured heavy drinking made it difficult to compare our results with those for other population samples, the other averages for the mean-age groups in each year of the given waves are roughly similar to national averages of alcohol consumption behavior (1, 19, 20).

Context

A person's alcohol use might mirror that of his or her social contacts.

Contribution

Using the same group of Framingham Heart Study participants who helped to define the associations between social networks and other health behaviors, the researchers found that alcohol use was similar among individuals in a social cluster. Furthermore, changes in a person's alcohol intake over time followed changes in the alcohol intake of their social contacts.

Caution

Alcohol use was self-reported, and the researchers did not have access to social contacts who were not participating in the study.

Implication

Changing alcohol use may require intervening with social groups as well as with individuals.

—The Editors

Statistical Analysis

Our first goal was to evaluate whether a person's alcohol consumption behavior was associated with that of his or her social network ties at various degrees of separation. To test this hypothesis, we took an observed clustering of persons (and their alcohol consumption behavior) within the whole network and compared them with 1000 simulated networks with the same topology and overall prevalence of drinking as the observed network, but with the incidence of drinking (for example, at least 1 drink per day) randomly distributed across the nodes ("random drinking networks"). If clustering occurs in drinking behavior, then the probability that a contact is a drinker given that a principal is a drinker should be higher in the observed network than in the random drinking networks (21). We used the Kamada–Kawai algorithm, which itera-

Table 1. **Summary Statistics for Principals**

Variables	Principals, %	Mean (SD)	Minimum	Lower Quartile	Upper Quartile	Maximum
All waves						
Continuous						
Drinks per day, *n*	–	0.88 (1.29)	0	0	1	17
Close friends, *n*	–	0.96 (0.88)	0	0	1	9
Family members, *n*	–	3.07 (3.59)	0	0	5	29
Contacts, *n*	–	2.70 (1.89)	1	1	4	19
Contacts who abstain, *n*	–	0.79 (1.02)	0	0	1	10
Contacts who drink heavily, *n*	–	0.56 (0.81)	0	0	1	7
Education, *y*	–	13.70 (2.29)	2	12	16	17
Age, *y*	–	50.87 (12.66)	21	42	60	90
Dichotomous, *n*						
Abstainers	29	–	–	–	–	–
Heavy drinkers	18	–	–	–	–	–
Women	52	–	–	–	–	–
Wave 1						
Continuous						
Drinks per day, *n*	–	1.06 (1.45)	0	0	1	14
Close friends, *n*	–	1.07 (0.84)	0	1	1	7
Family members, *n*	–	3.67 (3.96)	0	0	6	29
Contacts, *n*	–	3.11 (2.17)	1	1	4	17
Contacts who abstain, *n*	–	0.50 (0.80)	0	0	1	6
Contacts who drink heavily, *n*	–	0.76 (0.95)	0	0	1	6
Education, *y*	–	13.70 (2.29)	2	12	16	17
Age, *y*	–	38.06 (9.50)	21	30	45	70
Dichotomous, *n*						
Abstainers	15	–	–	–	–	–
Heavy drinkers	22	–	–	–	–	–
Women	52	–	–	–	–	–

tively repositions nodes to reduce the number of ties that cross each other, to draw the networks (22).

Our second goal was to examine the possible determinants of any clustering in alcohol consumption behavior. We considered 3 explanations for nonrandom clustering of alcohol consumption behavior in the network: principals might choose to associate with like contacts (homophily) (23, 24); principals and contacts might share attributes or jointly experience unobserved contemporaneous events that cause their alcohol consumption behavior to covary (omitted variables or confounding); and contacts might exert social influence or peer effects on principals (induction). The availability of dynamic, longitudinal data on both network connections and drinking behavior allowed us to distinguish between interpersonal induction of drinking and homophily (25).

Our basic statistical approach involved specifying longitudinal logistic regression models in which a principal's drinking status at time $t + 1$ is a function of his or her various attributes, such as age, sex, and education; his or her drinking status at time t; and the drinking status of his or her contacts at times t and $t + 1$ (25). We used generalized estimating equation procedures to account for multiple observations of the same principal across both waves and principal–contact pairings (26). We assumed an independent working correlation structure for the clusters (27).

By using a time-lagged dependent variable (lagged to the previous examination) for alcohol consumption, we

Table 2. **Average Age and Alcohol Consumption Behavior, by Examination**

Examination	Midpoint Year of Examination	Age, *y**	Drinks per Day, *n*	Abstainers, %	Heavy Drinkers, %†
1	1972	46.8	1.04	18.7	22.2
2	1981	53.0	0.99	30.1	21.8
3	1986	55.2	0.88	34.2	18.5
4	1989	57.5	0.76	35.8	15.6
5	1993	60.0	0.70	35.9	14.4
6	1997	63.1	0.63	42.5	12.7
7	2000	64.7	0.70	37.8	14.9

* Average age of principals across each examination wave.
† Defined as averaging more than 2 drinks per day for men and 1 drink per day for women.

eliminated serial correlation in the errors (28) (evaluated with a Lagrange multiplier test) and substantially controlled for the principal's genetic endowment and any intrinsic, stable predilection to drink. In addition, the lagged independent variable for a contact's drinking status substantially controlled for homophily (25, 29). The key variable of interest is a contact's alcohol consumption behavior at time $t + 1$. A significant coefficient on this variable would suggest either that a contact's drinking affects a principal's drinking or that a principal and a contact experience contemporaneous events that affect both of their alcohol consumption behaviors. We tested the possibility that omitted variables or unobserved events could explain the associations by examining how the type or direction of the social relationship between contacts affected the association between principal and contact drinking.

To calculate risk ratios and 95% CIs, we simulated the change in risk for principal drinking when contact contemporaneous drinking changes from 0 to 1 by using 1000 randomly drawn sets of estimates from the coefficient covariance matrix and assuming all other variables were held at their means (30). All of these tests are 2-tailed. For repeated tests that involved different types of social contacts, we applied a Bonferroni correction to the CIs.

We assessed the sensitivity of the results with multiple additional analyses (**Appendix**, available at www.annals.org). For example, we considered the possible effect of incomplete or biased network data. If people who drink heavily are more likely to name people outside the study, underestimation of the effect of one person's alcohol consumption behavior on another might occur. We found no significant correlation between number of drinks per day and number of ties to people outside the study ($\rho = 0.01$; $P = 0.15$), which suggests that the network data generation procedure did not bias the analyses.

Role of the Funding Source

The study was supported by the National Institutes of Health and the Robert Wood Johnson Foundation. The funding sources played no role in the design, conduct, or analysis of this study. Each researcher's institutional review board approved the study.

RESULTS

The **Appendix Figure**, available at www.annals.org, shows the largest connected subcomponent of the social network of friends, spouses, and siblings in the year 2000. Clusters of drinkers and abstainers can be seen in the network.

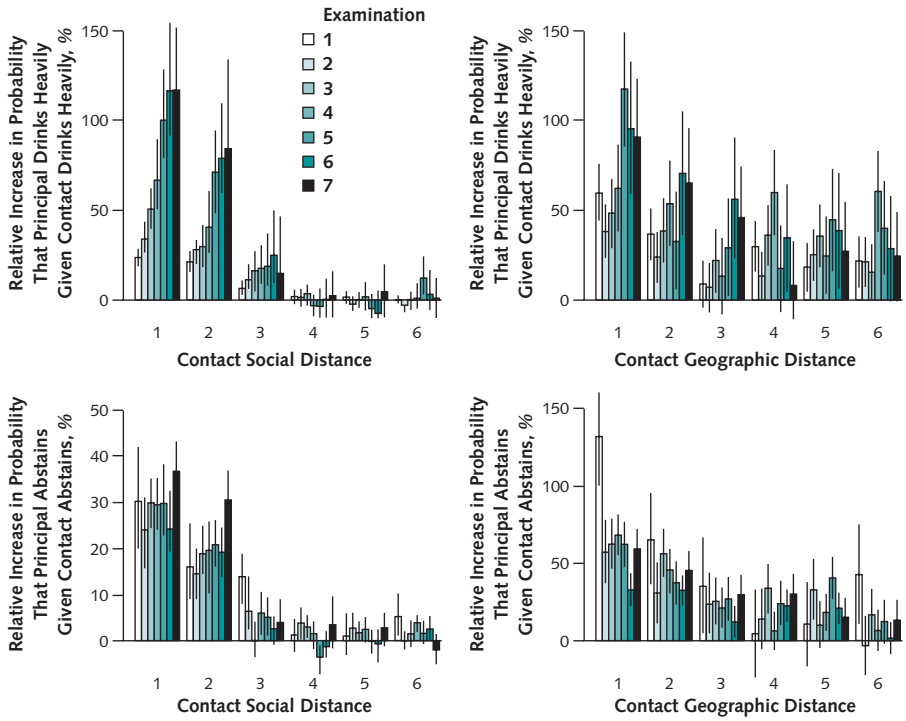
Figure 1 shows the correlation between principals and contacts with regard to their drinking behavior (the **Appendix**, available at www.annals.org, contains numerical results for this and the other figures). Our results indicate that principals are 50% (95% CI, 40% to 62%) more likely to drink heavily if a person they are directly connected to (1 degree of separation) drinks heavily. The size of the effect is 36% (CI, 25% to 48%) for people at 2 degrees of separation (such as the friend of a friend) and 15% (CI, 8% to 25%) for people at 3 degrees of separation (such as the friend of a friend of a friend). The effect disappears at 4 degrees of separation (4% [CI, −2% to 10%]), which is consistent with the "3 degrees of influence" rule of social network contagion that has been shown for obesity, smoking, happiness, depression, loneliness, word-of-mouth advertising, and the spread of ideas among inventors (10–14, 31). Analyses of the full network also show that persons are 29% (CI, 23% to 36%) more likely to abstain if someone they are directly connected to (1 degree of separation) abstains. The size of this effect is 21% (CI, 16% to 27%) at 2 degrees of separation and 5% (CI, 1% to 10%) at 3 degrees of separation, and it disappears at 4 degrees of separation (2% [CI, −1% to 6%]).

Of note, the decline in effect size with social distance in **Figure 2** contrasts with a lack of decline in effect size as people become more geographically distant from one another. We confirmed this by testing an interaction between distance and the effect size. Our results suggest that a friend or family member who lives hundreds of miles away is associated with as big an effect as one who lives next door.

The actual alcohol consumption behavior of social contacts affects a person's alcohol consumption behavior. **Figure 3** shows the smoothed bivariate relationship between the fraction of a principal's friends and family who drank heavily or abstained at one examination and the average number of drinks per day that principal reported at the next examination. Being surrounded by heavy drinkers increased the reported alcohol consumption by about 70% (CI, 35% to 142%) compared with those who were not connected to any heavy drinkers. Conversely, being surrounded by abstainers decreased reported alcohol consumption by half.

When we controlled for age, sex, education, and examination and regressed each principal's future alcohol consumption behavior on the basis of number of contacts who were heavy drinkers, moderate drinkers, or abstainers, we found that each additional heavy drinker increased the likelihood that a principal drank heavily by 18% (CI, 11% to 25%; $P < 0.001$) and decreased the likelihood that a principal abstained by 7% (CI, 2% to 12%; $P = 0.009$) but had no effect on moderate alcohol consumption behavior (CI, −8% to 1%; $P = 0.113$). Conversely, each additional abstainer significantly reduced the likelihood that a principal drank heavily by 10% (CI, 4% to 15%; $P = 0.001$), increased the likelihood that a principal abstained by 22% (CI, 17% to 28%), and decreased the likelihood that a principal drank moderately by 11% (CI, 8% to 14%). Finally, each additional moderate drinker had no significant effect on whether a principal drank heavily (CI, −2% to 7%; $P = 0.214$) but significantly decreased the probability that he or she abstained by 5% (CI, 2% to 9%) and increased the likelihood that he or she drank moderately by 6% (CI, 2% to 9%).

Figure 1. **Relationship of social and geographic distance to heavy drinking and abstaining in connected persons.**



We derived effects by comparing the conditional probability of drinking in the observed network with an identical network (with topology preserved) in which the same number of heavy drinkers or abstainers were randomly distributed. "Contact social distance" refers to the closest social distance (or degree of separation) between the contact and principal (for example, a direct friend = 1 and a friend's friend = 2). For geographic distance, we ranked all physical distances between the homes of directly connected principals and contacts (pairs at 1 degree of separation) and created 6 equally sized groups (1 = closest, 6 = farthest). The average distances for these groups are 0 miles (group 1), 0.26 mile (group 2), 1.5 miles (group 3), 3.4 miles (group 4), 9.3 miles (group 5), and 471 miles (group 6).

We next evaluated the extent of paired, interpersonal association in alcohol consumption behavior. As discussed, our models account for homophily by including a time-lagged measure of a contact's alcohol consumption behavior. We evaluated the possible role of unobserved contemporaneous events by separately analyzing models on subsets of the data that involved various principal–contact pairings. **Figure 3** summarizes the associations from the models (the **Appendix**, available at www.annals.org, contains numerical results). With respect to friends, we found significant sex differences in the spread of heavy alcohol consumption behavior. If a principal's female friend started drinking heavily, then the principal's chances of drinking heavily increased by 154% (CI, 30% to 354%). In contrast, a male friend's heavy alcohol consumption behavior seemed to have no significant effect on the principal. The type of friendship also seemed to be important: A woman did not seem to have a significant effect if she thought of the principal as a friend, but not vice versa (a contact-perceived friend), but the overlapping CIs indicate that the difference in the effect size is not significant. Sex also played a role among spouses. Heavy drinking by a wife increased the likelihood that the husband drank heavily by 196% (CI, 91% to 329%), whereas heavy drinking by a husband increased the likelihood that a wife drank heavily by 126% (CI, 67% to 202%). Among siblings, the effect was significantly smaller and did not differ whether the contact was a sister (37% [CI, 0% to 85%]) or a brother (34% [CI, 8% to 66%]). Immediate neighbors and coworkers had no significant effects on a principal's drinking behavior.

We repeated our analyses for abstention behavior and found broadly similar results. The effect of female friends abstaining was about the same size as that of male friends abstaining (42% [CI, 9% to 84%] vs. 44% [CI, −3% to 106%]), although the latter was barely insignificant. Wives who abstained seemed to have more influence than husbands (74% [CI, 40% to 115%] vs. 56% [CI, 32% to 82%]), but the effect of a sister was weaker than that of a brother (28% [CI, 13% to 45%] vs. 39% [CI, 19% to 60%]). Once again, immediate neighbors and coworkers had no effect on a principal's drinking behavior with respect to abstention.

DISCUSSION

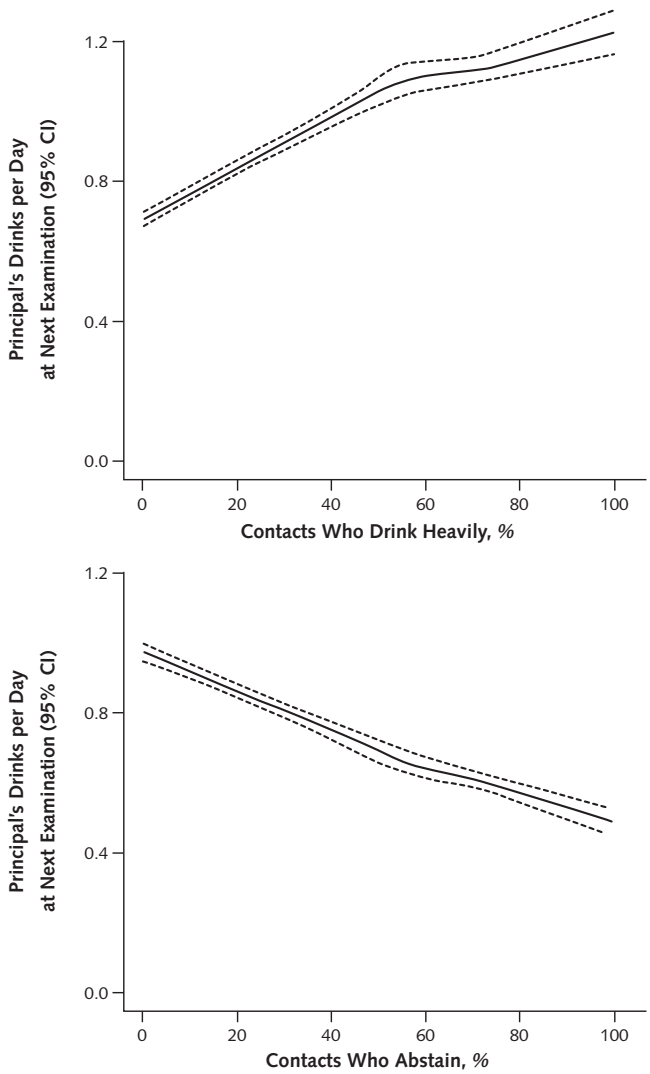
Alcohol consumption behavior among persons and those in their social networks is highly correlated. Interpersonal effects with respect to alcohol behavior vary in size according to the type of relationship. Induction of these

effects may occur through social norms (10, 12, 32–35); unfortunately, the study data include no measures of attitudes toward alcohol consumption, and claims about the underlying mechanisms for the network effects remain speculative.

Our general findings correspond with previous literature on obesity, smoking, happiness, and depression (10–14), although certain patterns of spread seem to be specific to alcohol use. One unique pattern we found relates to the bimodal nature of the social network effects. Whereas network effects were found for smoking cessation (11) (a positive health outcome) and for gaining weight (10) (a negative health outcome), the effect seems to be bidirectional in alcohol consumption with respect to both heavy drinking and abstaining. This suggests that social network effects may have both positive and negative health consequences for alcohol consumption behavior, depending on the circumstances.

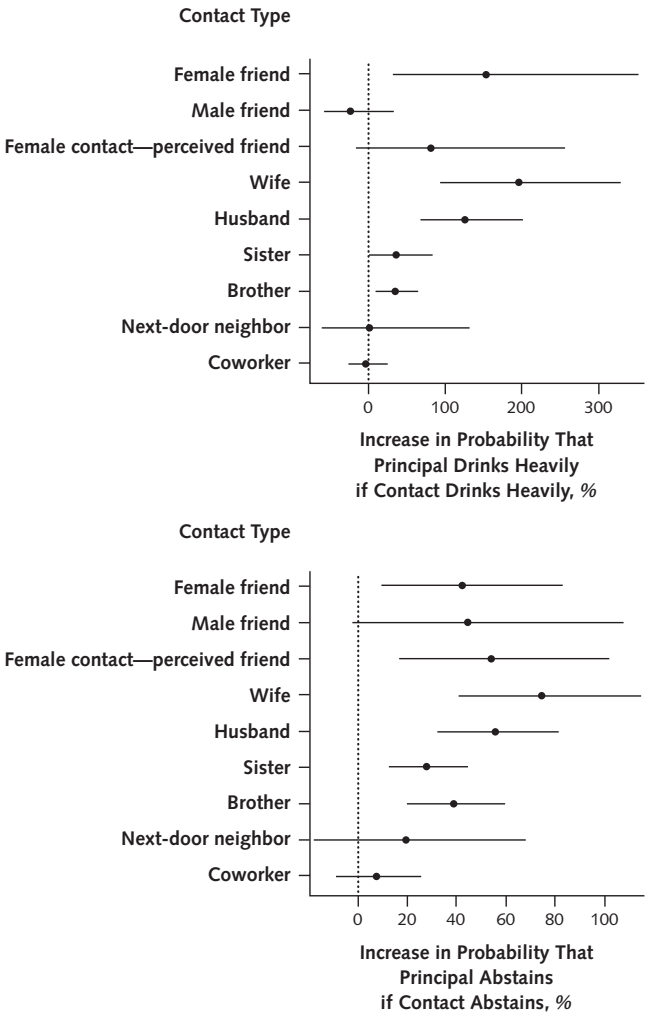
Another important finding relates to the role of sex in the spread of alcohol consumption behavior. Our findings suggest that female contacts are significantly more likely

Figure 2. **Effect of heavy drinking and abstaining contacts on principals in the Framingham Heart Study social network.**



Solid lines are based on bivariate LOESS regression, and dotted lines indicate 95% CIs. **Top.** Effect of contacts who drink heavily. **Bottom.** Effect of contacts who abstain.

Figure 3. **Contact type and drinking in the Framingham Heart Study social network.**



Changes in principal alcohol consumption given contact alcohol consumption are shown. Estimates are based on generalized estimating equation logit models of drinking in several subsamples of the Framingham Heart Study social network. The dependent variable in each model is principal drinking status. Independent variables include lagged principal drinking status; contact drinking status; lagged contact drinking status; principal age, sex, and education; and fixed effects for each wave. The **Appendix** (available at www.annals.org) contains full models and equations. To calculate mean effect sizes and Bonferroni-corrected 95% CIs, we simulated first difference in contact contemporaneous drinking (changing from 0 to 1) by using 1000 randomly drawn sets of estimates from the coefficient covariance matrix and assuming all other variables to be held at their means. **Top.** Effects of heavy drinking. **Bottom.** Effects of abstention.

than male contacts to influence the spread of heavy alcohol consumption behavior. Although differences may have been expected in principals of different sex (men and women perceiving peer influences and social norms about alcohol differently [36, 37]), the effect of contact sex was unexpected. One possible explanation is that significant increases in drinking behavior among women are much less common and more often associated with dramatic shifts in roles and contexts in life, such as job changes and work stress, which would reflect the effect of confounding factors (38). A related possibility is that changes in perceived norms toward drinking among women are more powerfully transmitted along social networks, possibly because women are usually perceived as sharing norms for less alcohol consumption (37, 38) and a woman who changes her behavior would therefore be a stronger stimulus.

Although our results have several significant associations, it is important to revisit whether they represent the spread of alcohol consumption behavior (induction) or reflect selection effects (homophily) or shared environmental effects (confounding) (23, 39, 40). Although we cannot completely rule out these alternative explanations, several of our findings strongly suggest that induction plays an important role (41). First, the directionality of friendship ties is significant in predicting the spread of alcohol consumption behavior. This provides some evidence for the interpersonal induction of alcohol consumption behavior and suggests that covariance in drinking between friends is not the result of mutual unobserved contemporaneous exposures. If it were, the influence should be equally strong regardless of the directionality of friendship. Second, our results show that neither immediate neighbors nor geographic distance modifies alcohol consumption behavior. If shared exposure (such as proximity to liquor stores or local economic hardship) were key, the effects would decay with distance. Third, because our models control for a principal's previous drinking status, we can account for sources of confounding that are stable over time (such as childhood exposures or genetic endowment). Finally, we can control for a contact's previous drinking status, thus accounting for a possible tendency of drinkers to form ties among themselves. To further control for homophily and environmental exposures, we are currently pursuing follow-up studies that use econometric and experimental methods.

Our study has limitations. First, our outcome measure is not a clinical tool, so we cannot make any specific conclusions about the spread of alcohol-related disorders per se in our sample. Second, we cannot estimate the relative negative health effect of increasing alcohol use, because alcohol use has been reported to have both positive and negative health effects. For example, moderate alcohol use is consistently associated with a lower risk for myocardial infarction (relative to abstention) in prospective cohort studies (42). This beneficial effect of moderate alcohol intake has been found to hold even for men with relatively healthy lifestyles (43, 44). In addition to cardiovascular effects, some evidence suggests that mild to moderate alcohol intake may be related to better cognitive functioning in older adults (45). Therefore, network effects that increase or decrease alcohol consumption could both have health benefits. Third, our sample is ethnically (but not socioeconomically) homogenous. Finally, all network ties were observed in the data set, which means our estimates may be biased.

Our results support the basic idea that because persons are connected, their health is also connected. Network phenomena might be exploited to spread positive health behaviors, a suggestion supported by numerous studies in the domain of drinking. For example, drinking cessation programs that provide peer support—that modify the social network of the target—are more successful (46–48). Of note, the oldest peer social support network in the country, Alcoholics Anonymous, is specifically designed to help foster social network connections, to encourage abstinence among its members and establish ties between principals and principal-identified contacts known as "sponsors." Alcoholics Anonymous reflects the creation of a kind of deliberate social network. Both good and bad behaviors may spread across a range of social ties at some distance from their origin. Our findings also reinforce the idea that drinking is a public health and clinical problem that involves groups of interconnected people who evince shared behaviors, and targeting these behaviors would rightly involve addressing groups and not just individuals.

From Massachusetts General Hospital, Harvard Medical School, and Boston University School of Medicine, Boston, Massachusetts; University of California, San Diego, San Diego, California; and Harvard University, Cambridge, Massachusetts.

Acknowledgment: The authors thank Laurie Meneades, Rebecca Joyce, Molly Collins, Marian Bellwood, and Karen Mutalik for the expert assistance required to build the data set.

Grant Support: By the National Institutes of Health (grant R01AG2444801) and a Pioneer Award from the Robert Wood Johnson Foundation. The Framingham Heart Study is supported by the National Heart, Lung, and Blood Institute (grant N01HC25195).

Potential Conflicts of Interest: Disclosures can be viewed at www.acponline.org/authors/icmje/ConflictOfInterestForms.do?msNum=M09-1111.

Reproducible Research Statement: *Study protocol and data set:* Not available. *Statistical code:* Available from Dr. Fowler (e-mail, jhfowler@ucsd.edu).

Requests for Single Reprints: Nicholas A. Christakis, MD, PhD, Harvard Medical School, 180-A Longwood Avenue, Boston, MA 02115; e-mail, christakis@hcp.med.harvard.edu.

Current author addresses and author contributions are available at www.annals.org.

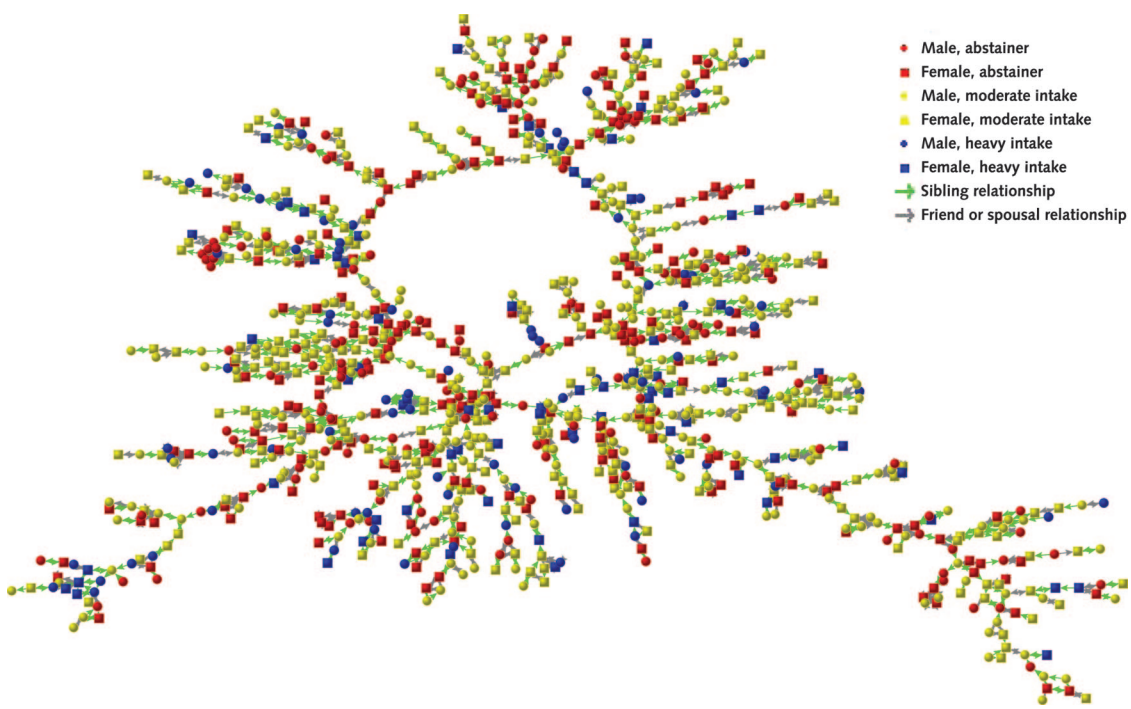
References

1. **Substance Abuse and Mental Health Services Administration.** Results from the 2002 National Survey on Drug Use and Health: National Findings, Office of Applied Studies, NHSDA 1999 Series H-22, DHHS Publication No. SMA 03–3836. Rockville, MD: Substance Abuse and Mental Health Services Administration; 2003.
2. **Naimi TS, Brewer RD, Mokdad A, Denny C, Serdula MK, Marks JS.** Binge drinking among US adults. JAMA. 2003;289:70-5. [PMID: 12503979]
3. **Kelley JF, Renner JA.** Alcohol-related disorders. In: Stern T, Rosenbaum J, Biederman J, Fava M, Rauch S, eds. The MGH Textbook of Comprehensive Clinical Psychiatry. Philadelphia: Mosby-Elsevier; 2008.
4. **Prescott CA, Kendler KS.** Genetic and environmental contributions to alcohol abuse and dependence in a population-based sample of male twins. Am J Psychiatry. 1999;156:34-40. [PMID: 9892295]
5. **Reifman A, Barnes GM, Dintcheff BA, Farrell MP, Uhteg L.** Parental and peer influences on the onset of heavier drinking among adolescents. J Stud Alcohol. 1998;59:311-7. [PMID: 9598712]
6. **Duncan GJ, Boisjoly J, Kremer M, Levy DM, Eccles J.** Peer effects in drug use and sex among college students. J Abnorm Child Psychol. 2005;33:375-85. [PMID: 15957564]
7. **Galea S, Nandi A, Vlahov D.** The social epidemiology of substance use. Epidemiol Rev. 2004;26:36-52. [PMID: 15234946]
8. **Pagan JL, Rose RJ, Viken RJ, Pulkkinen L, Kaprio J, Dick DM.** Genetic and environmental influences on stages of alcohol use across adolescence and into young adulthood. Behav Genet. 2006;36:483-97. [PMID: 16586152]
9. **Newman MEJ.** The structure and function of complex networks. SIAM Review. 2003;45:167-256.
10. **Christakis NA, Fowler JH.** The spread of obesity in a large social network over 32 years. N Engl J Med. 2007;357:370-9. [PMID: 17652652]
11. **Christakis NA, Fowler JH.** The collective dynamics of smoking in a large social network. N Engl J Med. 2008;358:2249-58. [PMID: 18499567]
12. **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. [PMID: 19056788]
13. **Christakis NA, Fowler JH.** Connected: The Suprising Power of Our Social Networks and How They Shape Our Lives. New York: Little Brown; 2009.
14. **Rosenquist JN, Christakis NA, Fowler JH.** The spread of depressive symptoms in a large social network over 32 years. Molecular Psychiatry. [In Press]
15. **Bearman PS, Moody J.** Suicide and friendships among American adolescents. Am J Public Health. 2004;94:89-95. [PMID: 14713704]
16. **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;110:281-90. [PMID: 474565]
17. **Fitzpatrick GL, Modlin ML.** Direct-Line Distances: International Edition. Metuchen, NJ: Scarecrow Pr; 1986.
18. **Del Boca FK, Darkes J.** The validity of self-reports of alcohol consumption: state of the science and challenges for research. Addiction. 2003;98 Suppl 2:1-12. [PMID: 14984237]
19. **Substance Abuse and Mental Health Services Administration.** 1998 National Survey on Drug Use and Health: Main Findings. Rockville, MD: Substance Abuse and Mental Health Services Administration; 1999.
20. **Substance Abuse and Mental Health Services Administration.** 2002 National Survey on Drug Use and Health: Main Findings. Rockville, MD: Substance Abuse and Mental Health Services Administration; 2003.
21. **Szabo G, Barabasi AL.** Network effects in service usage. arXiv.org; 2006. Accessed at http://lanl.arxiv.org/abs/physics/0611177 on 17 February 2010.
22. **Kamada T, Kawai S.** An algorithm for drawing general undirected graphs. Information Processing Letters. 1989;31:7-15.
23. **McPherson M, Smith-Lovin L, and Cook JM.** Birds of a feather: homophily in social networks. Annu Rev Sociol. 2001;27:415-444.
24. **Sackett DL, Anderson GD, Milner R, Feinleib M, Kannel WB.** Concordance for coronary risk factors among spouses. Circulation. 1975;52:589-95. [PMID: 1080450]
25. **Carrington PJ, Scott J, and Wasserman S.** Models and Methods in Social Network Analysis. New York: Cambridge Univ Pr; 2005.
26. **Liang KY, Zeger SL.** Longitudinal data analysis using generalized linear models. Biometrika 1986;73:13-22.
27. **Schildcrout JS, Heagerty PJ.** Regression analysis of longitudinal binary data with time-dependent environmental covariates: bias and efficiency. Biostatistics. 2005;6:633-52. [PMID: 15917376]
28. **Hosking JRM.** Lagrange-multiplier tests of time-series models. J R Statist Soc B 1980;42:170-181.
29. **Fowler JH, Christakis NA.** Estimating peer effects on health in social networks: a response to Cohen-Cole and Fletcher; and Trogdon, Nonnemaker, and Pais. J Health Econ. 2008;27:1400-5. [PMID: 18692263]
30. **Tanner, Martin A.** Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions. New York: Spring-Verlag; 1996.
31. **Cacioppo JT, Fowler JH, Christakis NA.** Alone in the crowd: the structure and spread of loneliness in a large social network. J Pers Soc Psychol. 2009;97:977-91. [PMID: 19968414]
32. **Berkowitz AD.** The Social Norms Approach: Theory, Research and Annotated Bibliography. Newton, MA: U.S. Department of Education; 2004.
33. **Borsari B, Carey KB.** Peer influences on college drinking: a review of the research. J Subst Abuse. 2001;13:391-424. [PMID: 11775073]
34. **Borsari B, Carey KB.** Descriptive and injunctive norms in college drinking: a meta-analytic integration. J Stud Alcohol. 2003;64:331-41. [PMID: 12817821]
35. **Perkins HW.** Social norms and the prevention of alcohol misuse in collegiate contexts. J Stud Alcohol Suppl. 2002:164-72. [PMID: 12022722]
36. **Prentice DA, Miller DT.** Pluralistic ignorance and the perpetuation of social norms by unwitting actors. In: Zanna MP, ed. Advances in Social Psychology. Vol. 28. San Diego, CA: Acad Pr. 1996.
37. **Wilsnack SC, Klassen AD, Schur BE, Wilsnack RW.** Predicting onset and chronicity of women's problem drinking: a five-year longitudinal analysis. Am J Public Health. 1991;81:305-18. [PMID: 1994739]
38. **Eagly AH, Wood W, Diekman AB.** Social role theory of sex differences and similarities: A current appraisal. In: Eckes T, Trautner HM, eds. The Development Social Psychology of Gender. Philadelphia: Lawrence Erlbaum; 2000.
39. **Bullers S, Cooper ML, Russell M.** Social network drinking and adult alcohol involvement: a longitudinal exploration of the direction of influence. Addict Behav. 2001;26:181-99. [PMID: 11316376]
40. **Peterson PL, Hawkins JD, Abbott RD, Catalano RF.** Disentangling the effects of parental drinking, family management, and parental alcohol norms on current drinking by black and white adolescents. J Res Adolesc. 1994;4:203-227
41. **Granovetter MS.** The strength of weak ties. Am J Sociol. 1973;78:1360-80.
42. **Maclure M.** Demonstration of deductive meta-analysis: ethanol intake and risk of myocardial infarction. Epidemiol Rev. 1993;15:328-51. [PMID: 8174661]
43. **Mukamal KJ, Chiuve SE, Rimm EB.** Alcohol consumption and risk for coronary heart disease in men with healthy lifestyles. Arch Intern Med. 2006;166:2145-50. [PMID: 17060546]
44. **Espeland MA, Coker LH, Wallace R, Rapp SR, Resnick SM, Limacher M, et al; Women's Health Initiative Study of Cognitive Aging.** Association between alcohol intake and domain-specific cognitive function in older women. Neuroepidemiology. 2006;27:1-12. [PMID: 16717476]
45. **Reid MC, Van Ness PH, Hawkins KA, Towle V, Concato J, Guo Z.** Light to moderate alcohol consumption is associated with better cognitive function among older male veterans receiving primary care. J Geriatr Psychiatry Neurol. 2006;19:98-105. [PMID: 16690995]
46. **Wing RR, Jeffery RW.** Benefits of recruiting participants with friends and increasing social support for weight loss and maintenance. J Consult Clin Psychol. 1999;67:132-8. [PMID: 10028217]
47. **Malchodi CS, Oncken C, Dornelas EA, Caramanica L, Gregonis E, Curry SL.** The effects of peer counseling on smoking cessation and reduction. Obstet Gynecol. 2003;101:504-10. [PMID: 12636954]
48. **McKnight AJ, McPherson K.** Evaluation of peer intervention training for high school alcohol safety education. Accid Anal Prev. 1986;18:339-47. [PMID: 3741584]

Current Author Addresses: Drs. Rosenquist and Christakis: Harvard Medical School, 180-A Longwood Avenue, Boston, MA 02115.
Dr. Murabito: Boston University School of Medicine, 73 Mount Wayte Avenue, Framingham, MA 01701.
Dr. Fowler: University of California, San Diego, Social Science Building 383, 9500 Gilman Drive, Number 0521, La Jolla, CA 92093-0521.

Author Contributions: Conception and design: J.H. Fowler, N.A. Christakis.
Analysis and interpretation of the data: J.N. Rosenquist, J. Murabito, J.H. Fowler, N.A. Christakis.
Drafting of the article: J.N. Rosenquist, J.H. Fowler, N.A. Christakis.
Critical revision of the article for important intellectual content: J.N. Rosenquist, J. Murabito, J.H. Fowler, N.A. Christakis.
Final approval of the article: J.N. Rosenquist, J.H. Fowler, N.A. Christakis.
Provision of study materials or patients: N.A. Christakis.
Statistical expertise: J.H. Fowler, N.A. Christakis.
Obtaining of funding: J.H. Fowler, N.A. Christakis.
Administrative, technical, or logistic support: J.H. Fowler, N.A. Christakis.
Collection and assembly of data: J. Murabito, J.H. Fowler, N.A. Christakis.

Appendix Figure. **Drinking in the Framingham Heart Study social network in 2000.**



A sample of the largest component of friends, spouses, and siblings at examination 7 (centered on the year 2000); 1073 individuals are shown. Each node represents 1 person. The graph suggests clustering in abstention and heavy alcohol consumption behavior, both of which are confirmed by statistical models.

Candidate Performance Measures for Screening for, Assessing, and Treating Unhealthy Substance Use in Hospitals: Advocacy or Evidence-Based Practice?

Richard Saitz, MD, MPH

The Joint Commission recently proposed candidate performance measures addressing unhealthy substance use in hospitalized patients. The proposed measures of screening and brief intervention (SBI) assume that interventions that work in one setting (primary care outpatient practice) would work in another (hospital); treatment would have the same benefits for persons identified by screening as for those with symptoms who seek help; treatments that work for persons less severely affected by substance use would also work for those with more severe illness; and an approach that works for nondependent, unhealthy alcohol use would work for drug use. However, these assumptions extrapolate evidence of the effectiveness of SBI for primary care outpatients with nondependent, unhealthy alcohol use to the inpatient setting, persons with dependence, and other substances. Although quality of care for unhealthy substance use in all medical settings needs to improve, the evidence base for SBI in the hospital is too limited for the implementation of performance measures assessing this care.

Ann Intern Med. 2010;153:40-43. www.annals.org
For author affiliation, see end of text.

Unhealthy substance use is the spectrum from risky alcohol and other drug use (for example, marijuana, stimulants, opioids) through dependence. A technical advisory panel for the Joint Commission recently proposed candidate performance measures for evaluating screening and brief intervention (SBI) for unhealthy substance use in hospitalized patients (**Table**) (1). The proposed measures raise questions about the balance between advocacy and evidence when considering policy.

Persons with unhealthy substance use in medical settings are often unrecognized and do not receive minimally acceptable management. Performance measurement makes sense as part of a strategy to improve care; however, performance measures must be evidence-based (2), and the proposed practices are not.

The proposed measures have some technical limitations that could be solved when final measures are developed. For example, measures should be separate for alcohol and drug use because SBI tools and practices differ and because specificity would provide information about which practices to improve. In addition, the measures should specify the frequency of screening. Reporting would not be possible without that specification, but the conceptual problem with the measures is of greater interest and is worth broader discussion.

Several assumptions have been made in the development of these measures: interventions that work in one setting (primary care) would work in another (hospital); treatment would have the same benefits for persons identified by screening as for those with symptoms who seek help; treatments that work for persons less severely affected would work for those with more severe illness; and an approach that works for one condition would work for another similar but distinct condition. Randomized trials do not support these assumptions.

On the basis of 12 randomized trials in primary care settings, the U.S. Preventive Services Task Force recommended SBI for nondependent, unhealthy alcohol (but not other drug) use in outpatient primary care settings (3, 4). These studies have shown modest decreases in alcohol consumption at 1 year (4). The proposed Joint Commission measures (1) cite this evidence as supporting alcohol and drug SBI in inpatients, yet these results may not translate there. The best evidence for SBI efficacy is in primary care (4), where screening is often done by a patient's clinician and in a context the patient knows and visits longitudinally for preventive and comprehensive care.

Screening and brief intervention for drugs and alcohol is a cornerstone of the Office of National Drug Control Policy's strategy (5). The Substance Abuse and Mental Health Services Administration has spent several hundred million dollars on SBI in diverse general health settings, including emergency departments, trauma and primary care centers, and hospitals. A pre–post, uncontrolled, retrospective evaluation of self-reported data collected to meet government administrative requirements reported dramatic decreases in drug use after brief intervention in those settings (6). Among patients identified as eligible for brief intervention, treatment, or referral at 6 sites, 10% were selected for follow-up, and 4% to 75% of the selected patients across those sites were lost. This evidence was cited to support the proposed hospital measures (1), but the effect sizes observed are not plausible when one considers the modest effects seen in meta-analysis of controlled trials of SBI for alcohol use in primary care (4). Secular trend and self-change are more likely explanations, with some smaller contribution made by regression to the mean.

See also:

Web-Only
Conversion of graphics into slides

Table. **The Joint Commission Candidate Performance Measures for Assessing and Treating Alcohol and Other Drug Use and Dependence***

Name of Measure	Proposed Numerator	Proposed Denominator
Alcohol and other drug use screening	The number of hospitalized inpatients 12 y or older who were screened for alcohol and drug use by using validated screening tools for unhealthy drinking and drug use	The number of hospitalized inpatients 12 y or older
Alcohol and other drug use and dependence—brief intervention or treatment	The number of hospitalized inpatients 12 y or older who were identified with unhealthy alcohol or drug use and received a brief intervention or, if found to have dependence, received substance use treatment as an inpatient or a referral for treatment	The number of hospitalized inpatients 12 y or older who were identified with unhealthy or dependent alcohol or drug use, including prescription drug misuse
Alcohol and other drug use and dependence—treatment management at discharge	The number of hospitalized inpatients 12 y or older who were identified as dependent on alcohol or drugs who 1) received/filled at discharge a prescription for medication for alcohol or opioid dependence with follow-up appointment(s) and 2) received a referral to an addictions treatment program, addiction treatment specialist, or mental health program or mental health specialist explicitly for follow-up for substance use or addiction treatment or 3) received a referral to a medical or health professional explicitly for follow-up for substance use or addiction	The number of hospitalized inpatients 12 y or older who were identified with alcohol or drug dependence or abuse disorders
Alcohol and other drug use and dependence—follow-up for unhealthy use and/or disorders	The number of discharged patients who were identified during a hospital stay with unhealthy alcohol or drug use or alcohol and/or drug use disorders who indicated they decreased or stopped using alcohol or drugs or have received follow-up addiction treatment from a specialty substance use treatment program, behavioral health specialist, or medical professional	The number of discharged patients 12 y or older who screened positive for unhealthy alcohol or drug use or dependence who received a follow-up call within 2 wk after hospital discharge

* Candidate measures were posted for public comment in September 2009 at www.jointcommission.org/NR/rdonlyres/05904E01-E5F7-49F1-8AD2-EC5058D91794/0/CandidateFinalMIFTADD8.doc. Tobacco measures were included in that posting but are not shown because they are not the subject of this article. A notice accompanying the candidate measures states, "All candidate measures are subject to change and may be eliminated from future consideration."

To my knowledge, only 2 randomized trials of drug SBI, in outpatient adults have been published. One reason for this is the paucity of brief drug-screening tools validated in general health settings (7). The relatively low prevalence of drug use is probably another reason. One small published study was in Brazil, and it deserves replication (8). The other was a large outpatient study (urgent care, homeless, and women's clinic) that showed modest benefits: Abstinence was greater by 9% for opioids and 5% for cocaine compared with control participants (9). The National Institute on Drug Abuse and the Substance Abuse and Mental Health Services Administration have funded trials of drug SBI in primary care, recognizing that the evidence is insufficient, as has the U.S. Preventive Services Task Force (2). To my knowledge, no randomized trials of SBI for the spectrum of other drug use in inpatients have been published.

Although cautious optimism may be warranted on the basis of the 2 outpatient studies, we have good reason for concern that brief intervention after drug screening will not be efficacious. One reason is that the target condition is heterogeneous. It is at least plausible that persons identified by screening with occasional marijuana use will have a different response to brief intervention than persons who inject heroin many times daily. Another reason for concern is that the proportion of persons identified by drug screening who have dependence is higher than that of those with unhealthy alcohol use identified by screening. Dependence is probably not as responsive to brief intervention as nondependent use.

We also have good reason for concern that alcohol or drug SBI will not be efficacious in the hospital. The prevalence of dependence is much higher among persons identified by screening in the hospital than in primary care settings. Brief intervention in persons identified by screening has proven efficacy only for those without dependence. Brief intervention could "work" for persons with dependence even if it did not reduce use or consequences if it led patients to enter addiction treatment. However, few studies have addressed this outcome (10). In 1 trial, brief intervention substantially increased linkage to alcohol treatment, but participants with dependence were largely excluded, even though most inpatients with unhealthy alcohol use have dependence (11). A second trial reported no increase in completed referrals for alcohol treatment (12).

A systematic review identified 11 alcohol SBI studies of 2441 patients who received medical, surgical, orthopedic, and trauma inpatient services (10). Results were inconclusive. Brief intervention reduced weekly alcohol consumption at 6 months in 3 studies with substantial heterogeneity. Brief intervention did not reduce weekly drinking at 1 year, laboratory markers, self-report of heavy drinking or driving offenses, or death. Exclusion of 1 of the studies that tested a not-so-brief 3-session intervention and had nonblinded outcome assessments made the 6-month weekly consumption results nonsignificant.

Of interest, alcohol screening is now a requirement for accreditation of trauma centers (13). This standard was largely based on 1 single-site study that had 54% follow-up for the self-report consumption outcome that improved with brief intervention. The study found a nonsignificant reduction in reinjury in administrative data for the whole sample (hazard ratio, 0.52 [95% CI, 0.21 to 1.29]) (14). Four other randomized trials inform the question of efficacy of alcohol SBI in hospitalized trauma patients. In a study of 126 patients, a brief intervention-associated decrease in driving while intoxicated was not significant (15). The decrease became significant in multivariable analyses despite the absence of baseline differences between groups. Another study of 120 trauma patients was not an alcohol SBI study per se; patients were not identified by alcohol screening, nor did they have to have unhealthy alcohol use (16). A collaborative care intervention that addressed posttraumatic stress disorder, depressive symptoms, and alcohol use prevented a 1-year increase in the proportion of patients with an alcohol use disorder; no differences at 6 months occurred, and drinking was not an outcome. Results from the 2 other studies were negative (17, 18). These 5 studies are not conclusive about the effect of universal alcohol SBI in trauma centers: 1 had limited follow-up and a nonsignificant main result; 1 was small, with positive results only in a secondary analysis; 1 showed benefits of a disease management intervention; and 2 had negative results.

The proposed measures go beyond SBI to address treatment and follow-up (**Table**). Treatment and follow-up for persons with recognized substance dependence make sense if patients are ready and willing, particularly for those who ask for help, but giving it to all persons identified by screening is of unknown effectiveness and may be counterproductive if patients perceive that they are being pushed into treatment. Behavior-change counseling can help (4), but effectiveness probably varies by context (10).

The Joint Commission should be commended for trying to address the poor quality of care that patients with unhealthy substance use receive in hospitals, but performance measures for universal alcohol and drug SBI in hospitals risk misdirecting efforts to ineffective practices and may divert attention from currently fixable problems.

Screening and brief intervention for inpatients seem to be simple, logical, and inexpensive. Hospitalization could be a teachable moment—and SBI dissemination could highlight the problem—and in theory improve care. These reasons could justify demonstration projects but not performance measures. Performance measures should assess evidence-based practices, and SBI in hospitals is not evidence-based. In addition, SBI is not simple, inexpensive, or obviously effective. Adding validated questionnaires to busy hospital routines is not simple, particularly because of the need to address positive results. Many institutions have hired staff to implement SBI (6). These complexities and costs and the potential harms of breaches of confidentiality, stigma, discrimination, and unpredictable effects on substance use could be justifiable, but only if hospital SBI has proven health benefits.

Patients with addictions who are in hospitals and seek help have great difficulty receiving addiction-specialty treatment. The first step toward improving quality of care should be taking care of these patients. Performance measures could address evidence-based care, such as pharmacotherapy for opioid and alcohol dependence and counseling. Hospitals should also have addiction treatment referral resources available for care after discharge. Routine practice should include asking about and documenting substance use, as clinicians do for medications. The purpose of identifying substance use is to consider it when making medical diagnoses and implementing treatments, such as prescriptions for addictive medications for treating pain. However, asking should not be confused with SBI, the use of a validated questionnaire, and counseling aimed at substance use.

In primary care, performance measures (19) and other strategies should be implemented to improve low rates of alcohol SBI. The Joint Commission measures could be developed for alcohol SBI in primary care settings that would be evidence-based. Research and demonstration projects should begin to address the growing problem of prescription drug abuse and the integration of behavioral health issues, including substance use and medical care. However, the evidence is insufficient to develop and implement performance measures for practices that may not improve outcomes. Unhealthy substance use is common and the cause of suffering for many. Although advocacy and policy play very important roles in getting good care to patients, this is not the time to allow them to trump the need for evidence to support best practices for patients.

From Boston Medical Center, Boston University School of Medicine, and Boston University School of Public Health, Boston, Massachusetts.

Acknowledgment: The author thanks Dr. Jeffrey H. Samet and anonymous peer reviewers who critically reviewed the manuscript, which led to substantial improvements.

Potential Conflicts of Interest: Disclosures can be viewed at www.acponline.org/authors/icmje/ConflictOfInterestForms.do?msNum=M09-2561.

Requests for Single Reprints: Richard Saitz, MD, MPH, Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, Boston Medical Center, 801 Massachusetts Avenue, 2nd Floor, Boston, MA 02118; e-mail, rsaitz@bu.edu.

References

1. **The Joint Commission.** Candidate Measure Profile for Assessing and Treating Tobacco, Alcohol, and Other Drug Use and Dependence. Accessed at www.jointcommission.org/NR/rdonlyres/05904E01-E5F7-49F1-8AD2-EC5058D91794/0/CandidateFinalMIFTADD8.doc on 29 September 2009.

2. **The Joint Commission.** Attributes of Core Performance Measures and Associated Evaluation Criteria. Accessed at www.jointcommission.org/NR/rdonlyres/7DF24897-A700-4013-A0BD-154881FB2321/0/AttributesofCorePerformance

MeasuresandAssociatedEvaluationCriteria.pdf on 24 January 2010.
3. U.S. Preventive Services Task Force. Accessed at www.ahrq.gov/CLINIC/uspstfix.htm on 28 October 2009.
4. **Whitlock EP, Polen MR, Green CA, Orleans T, Klein J; U.S. Preventive Services Task Force.** Behavioral counseling interventions in primary care to reduce risky/harmful alcohol use by adults: a summary of the evidence for the U.S. Preventive Services Task Force. Ann Intern Med. 2004;140:557-68. [PMID: 15068985]
5. National Drug Control Strategy 2010. Accessed at www.whitehousedrugpolicy.gov/publications/policy/ndcs10/ndcs2010.pdf on 17 May 2010.
6. **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;99:280-95. [PMID: 18929451]
7. **Smith PC, Schmidt SM, Allensworth-Davies D, Saitz R.** A single-question screening test for drug use in primary care. Arch Intern Med. 2010. [Forthcoming].
8. **De Micheli D, Fisberg M, Formigoni ML.** [Study on the effectiveness of brief intervention for alcohol and other drug use directed to adolescents in a primary health care unit]. Rev Assoc Med Bras. 2004;50:305-13. [PMID: 15499485]
9. **Bernstein J, Bernstein E, Tassiopoulos K, Heeren T, Levenson S, Hingson R.** Brief motivational intervention at a clinic visit reduces cocaine and heroin use. Drug Alcohol Depend. 2005;77:49-59. [PMID: 15607841]
10. **McQueen J, Howe TE, Allan L, Mains D.** Brief interventions for heavy alcohol users admitted to general hospital wards. Cochrane Database Syst Rev. 2009:CD005191. [PMID: 19588369]
11. **Elvy GA, Wells JE, Baird KA.** Attempted referral as intervention for problem drinking in the general hospital. Br J Addict. 1988;83:83-9. [PMID: 3345386]
12. **Saitz R, Palfai TP, Cheng DM, Horton NJ, Freedner N, Dukes K, et al.** Brief intervention for medical inpatients with unhealthy alcohol use: a randomized, controlled trial. Ann Intern Med. 2007;146:167-76. [PMID: 17283347]
13. **Committee on Trauma, American College of Surgeons.** FAQ for Resources for Optimal Care of the Injured Patient: 2006. Accessed at www.facs.org/trauma/faq_answers.html on 28 October 2009.
14. **Gentilello LM, Rivara FP, Donovan DM, Jurkovich GJ, Daranciang E, Dunn CW, et al.** Alcohol interventions in a trauma center as a means of reducing the risk of injury recurrence. Ann Surg. 1999;230:473-80. [PMID: 10522717]
15. **Schermer CR, Moyers TB, Miller WR, Bloomfield LA.** Trauma center brief interventions for alcohol disorders decrease subsequent driving under the influence arrests. J Trauma. 2006;60:29-34. [PMID: 16456433]
16. **Zatzick D, Roy-Byrne P, Russo J, Rivara F, Droesch R, Wagner A, et al.** A randomized effectiveness trial of stepped collaborative care for acutely injured trauma survivors. Arch Gen Psychiatry. 2004;61:498-506. [PMID: 15123495]
17. **Sommers MS, Dyehouse JM, Howe SR, Fleming M, Fargo JD, Schafer JC.** Effectiveness of brief interventions after alcohol-related vehicular injury: A randomized controlled trial. J Trauma. 2006;61:523-31. [PMID: 16966982]
18. **Soderstrom CA, DiClemente CC, Dischinger PC, Hebel JR, McDuff DR, Auman KM, et al.** A controlled trial of brief intervention versus brief advice for at-risk drinking trauma center patients. J Trauma. 2007;62:1102-11. [PMID: 17495708]
19. **Centers for Medicare & Medicaid Services.** Physician Quality Reporting Initiative. Accessed at www.cms.hhs.gov/pqri/ on 28 October 2009.

Correction

The first sentence of the penultimate paragraph of the Perspective by Saitz (1) should read as follows: "Patients with addictions who are in hospitals and seek help have great difficulty receiving addiction-specialty treatment."

The online version has been corrected.

Reference

1. Saitz R. Candidate performance measures for screening for, assessing, and treating unhealthy substance use in hospitals: advocacy or evidence-based practice? Ann Intern Med. 2010;153:40-3.

Drug and Alcohol REVIEW

APSAD

Drug and Alcohol Review (November 2010), 29, 631–640
DOI: 10.1111/j.1465-3362.2010.00217.x

Alcohol screening and brief intervention in primary care: Absence of evidence for efficacy in people with dependence or very heavy drinking

RICHARD SAITZ

Boston University Schools of Medicine and Public Health and Boston Medical Center, Boston, USA

Abstract

Issues. *Although screening and brief intervention (BI) in the primary-care setting reduces unhealthy alcohol use, its efficacy among patients with dependence has not been established. This systematic review sought to determine whether evidence exists for BI efficacy among patients with alcohol dependence identified by screening in primary-care settings.* ***Approach.*** *We included randomised controlled trials (RCTs) extracted from eight systematic reviews and electronic database searches published through September 2009. These RCTs compared outcomes among adults with unhealthy alcohol use identified by screening who received BI in a primary-care setting with those who received no intervention.* ***Key Findings.*** *Sixteen RCTs, including 6839 patients, met the inclusion criteria. Of these, 14 excluded some or all persons with very heavy alcohol use or dependence; one in which 35% of 175 patients had dependence found no difference in an alcohol severity score between groups; and one in which 58% of 24 female patients had dependence showed no efficacy.* ***Conclusion and Implications.*** *Alcohol screening and BI has efficacy in primary care for patients with unhealthy alcohol use, but there is no evidence for efficacy among those with very heavy use or dependence. As alcohol screening identifies both dependent and non-dependent unhealthy use, the absence of evidence for the efficacy of BI among primary-care patients with screening-identified alcohol dependence raises questions regarding the efficiency of screening and BI, particularly in settings where dependence is common. The finding also highlights the need to develop new approaches to help such patients, particularly if screening and BI are to be disseminated widely.* [Saitz R. Alcohol screening and brief intervention in primary care: Absence of evidence for efficacy in people with dependence or very heavy drinking. *Drug Alcohol Rev* 2010;29;631–640]

Key words: alcohol, alcohol dependence, primary care, brief intervention, systematic review.

Background

Alcohol brief intervention (BI) has proven efficacy in primary-care patients with non-dependent unhealthy alcohol use identified by screening [1–4]. Systematic reviews of controlled trials have found a reduction in alcohol consumption of 38 g per week [2] and a 12% reduction in the proportion of patients drinking risky amounts [1] among patients receiving BI compared with those receiving no intervention. However, screening identifies people with the range of unhealthy alcohol use, from risky use without consequences through dependence, and the efficacy of BI for patients with very heavy use or dependence, particularly those identified by screening, has not been established. This observation is important, as 20% of primary-care patients with unhealthy alcohol use identified by screening have dependence [5].

Furthermore, readiness to change and effectiveness of treatment should not be assumed to be the same for people seeking help versus those identified by screening. People with non-dependent unhealthy use may simply be unaware of the consumption amounts that risk health consequences or may have experienced few, if any, consequences related to drinking. As a result, they often are not seeking help or advice. In that circumstance, and without loss of control over their drinking, one would expect them to respond to brief advice from their physician, whom they see for preventive and primary care, when it is offered appropriately.

Richard Saitz MD, MPH, FACP, FASAM, Professor of Medicine and Epidemiology. Correspondence to Dr Richard Saitz, Boston Medical Center, 801 Massachusetts Avenue, 2nd Floor, Boston, MA 02118, USA. Tel: +1 617 414 7744; Fax: +1 617 414 4676; E-mail: rsaitz@bu.edu

Received 24 December 2009; accepted for publication 2 June 2010.

On the other hand, people with dependence who seek help often *do* benefit from treatment, including BI [6]. In such cases, the patient has already taken the first steps towards change, unlike the patient identified by screening. Although patients with dependence often report high readiness to change [7], by definition (i.e. dependence criteria), they have great difficulty doing so. Those identified through screening are not actively seeking help and therefore are less likely to be amenable to change. Thus, the literature on efficacy of treatment for people with alcohol dependence may not apply to those identified by screening. The widespread and continued existence of (not brief) specialty care for alcoholism confirms that experts have not concluded that BI is sufficient treatment for dependence. Furthermore, some believe that severity of unhealthy use is an explanation for negative studies of alcohol BI [8].

In clinical practice, the severity of alcohol use is not known until after screening, and clinicians need something efficacious to offer all patients identified as having unhealthy use, including those with greater severity or with dependence. The solution has been to perform BI among patients with dependence with a goal of motivating change, including referral and treatment. Yet it remains unknown whether BI in such patients leads to a decrease in use or alcohol-related consequences or to linkage with treatment. This systematic review of randomised controlled trials (RCTs) sought to find evidence to determine whether BI among patients identified by screening in primary-care settings as having alcohol dependence decreases consumption or alcohol-related consequences or increases initiation of, or engagement in, further alcohol treatment.

Methods

Inclusion criteria

This analysis included RCTs published in English in the peer-reviewed literature through September 2009 that compared primary-care patients with unhealthy alcohol use identified by screening who received BI with patients who received no intervention. BIs were conducted in-person (not by telephone, mail or computer). Each could include up to four follow-up sessions. The goal of the BI could be to reduce drinking and/or alcohol consequences or to provide a referral to additional care.

Exclusion criteria

Studies conducted among hospital inpatients or in emergency departments, trauma centres or other settings not considered primary care per the US Institute of Medicine definition (i.e. integrated, accessible health care by clinicians who are accountable for addressing a large majority of personal health-care needs, who develop a sustained partnership with patients, and who practice in the context of family and community [9]) were excluded, as were studies including patients with comorbid conditions (e.g. gastrointestinal disease, hypertension or pregnancy). Studies that compared BI with another active treatment (vs. usual care or no intervention) were also excluded, as were studies among patients not identified by screening, as motivation for change and severity of use can be substantially different in such patients. Additional exclusions were duplicate reports of results from the same study and studies with methodological flaws as defined in Whitlock *et al.* [4]. They used internal-validity [10] and quality [11] criteria to exclude trials of poor quality. Major quality problems were non-random assignment, non-comparable baseline conditions, attrition greater than 30%, inadequate or unavailable consumption data, or lack of data on alcohol-related consequences or treatment linkage outcomes [4].

Search strategy

Studies were selected from two recent exhaustive, high-quality systematic reviews by Kaner *et al.* [3] and Whitlock *et al.* [4], which identified RCTs of alcohol BI among primary-care patients through 2006. References from six other systematic reviews of alcohol BI in primary care were examined to identify additional studies [1,2,12–15]. Other systematic reviews may exist but were not identified and therefore were not included in this review. Finally, an electronic literature search was conducted to identify RCTs published from 2006 through September 2009 using an inclusive search strategy that combined relevant keywords and medical subject headings (MeSH) across five relevant and comprehensive online databases (Appendix 1).

Study selection

Thirty-three studies were identified by Whitlock *et al.* [4] and Kaner *et al.* [3]. Of these, 19 were excluded from this analysis, the reasons for which are listed in Appendix 2. The remaining 14 studies [16–29] were included in this analysis. Examination of the six additional systematic reviews [1,2,12–15] identified one RCT [30] meeting inclusion criteria (found in Bertholet *et al.* [2]), which was also included in this analysis.

The electronic search of Medline and the Cochrane Database of Systematic Reviews identified 227 potential articles published between 2006 and September 2009, of which eight were studies of alcohol BI. Seven

of the eight studies did not meet inclusion criteria. The reasons for exclusion are listed in Appendix 2. Searches of four additional relevant databases yielded a number of potential additions, but no additional studies meeting inclusion criteria. Thus, one study [31] was included from this search. Combined with the RCTs identified in the previously published systematic reviews, a total of 16 studies (N = 6839 patients) were included in this analysis (Table 1).

Findings

Of the 16 RCTs meeting inclusion criteria, 14 excluded some or all subjects with very heavy alcohol use or dependence, the definitions of which were specific to each study (Table 2). Only two studies included patients with dependence or did not exclude people based on an upper limit of consumption (Table 3). In the study by Burge *et al.* [30], 10–15 min BI by resident family physicians was compared with six weekly 90 min educational sessions, receipt of both interventions or receipt of no intervention among 175 Mexican Americans (75% of whom were men, 35% of whom had dependence and 65% of whom had abuse). Follow up occurred over 18 months. No difference was found between groups on drinks per week or Addiction Severity Index (ASI) alcohol scores. An interaction between BI and ASI alcohol score was not significant, suggesting a similar lack of response to intervention across the range of severity. The ASI family scale score improved among all subjects at 12 months, but at 18 months, BI was associated with a loss of this initial improvement among women. All groups showed improvement in the ASI medical score at 12 months, but only those in either intervention group continued to improve at 18 months, while those who received both interventions or no intervention did not. There were no group outcome differences in employment, legal problem severity, psychiatric severity or laboratory test results (mean corpuscular volume, gamma-glutamyltransferase or alanine or aspartate aminotransferase levels). In the study by Chang *et al.* [17], including 24 women (58% of whom had dependence, 8% of whom had abuse and 21% of whom had a past alcohol use disorder), no difference in alcohol consumption was observed between groups, the BI was done by an experienced addiction psychiatrist, and duration was not specified. Both study samples were recruited from single clinic sites. Neither study addressed treatment linkage as a goal of BI, although in Chang *et al.* referral to treatment was the control condition. Despite this, no control patients followed through on the referral in that study, and linkage to treatment was not reported for the BI group.

Conclusions

Results of the two studies identified in this systematic review offer no evidence to support alcohol screening and BI efficacy among primary-care patients with very heavy drinking or dependence. Furthermore, the studies were of limited generalisability. Patient samples were small in both studies; one study included only women, and the other included only Mexican Americans; results were not specifically analysed by dependence; the range of outcome measures could have been greater; and, in one of the studies, the intervention was conducted by an expert.

Other evidence in the literature suggests that screening and BI has efficacy for those who drink too much but do not have dependence. It may turn out to have efficacy for those with more severe unhealthy use, but evidence to date is not available to support this. There is a difference between an absence of evidence and evidence for absence of effect: the circumstance here is the former. Regardless, it is clear that we cannot conclude, on the basis of high-quality evidence in the scientific literature, that BI among those identified by alcohol screening in primary care works for people with very heavy drinking or dependence.

This study had several limitations. Ideally, systematic reviews are conducted by at least two independent raters, and interrater reliability is reported. Such a process might have strengthened this paper. Second, the findings relied largely on prior systematic reviews, exclusion criteria in original articles were sometimes not clearly specified, and non-English language studies were not included. However, of the four non-English studies included in the prior systematic reviews, two excluded patients with dependence and patients who had received treatment for alcohol problems [32,33], one excluded patients who had received treatment [34], and one excluded people with heavy drinking (≥95 units per week) and failed to provide clear data on randomisation methods [35]. As such, these studies would have either been excluded from the current review based on criteria other than language or, if included, would have appeared among the studies that excluded patients with very heavy use or dependence (and thus would not have contributed information on efficacy of BI in such patients).

One might wonder whether the BI literature beyond primary care might inform the question asked in this review. However, the context of care appears to be quite important for BI. In the literature on BI in emergency departments, most studies found no impact of BI on drinking [36–38]. Severity may be an explanation for inconsistent results on BI efficacy in other settings, such as hospitals [39]. For example, one of the few BI studies in any setting that did not exclude people with

Table 1. *Summary of alcohol screening and BI trials included in the systematic review*

Author (year)	Setting	Participants (age)	Intervention
Anderson & Scott (1992) [16]	8 primary-care group practices in England	154 men (17–69 years)	10 min BI, including advice and feedback from usual physician provider
Burge *et al.* (1997) [30]	US family practice outpatient clinic	175 Mexican-American adult men and women (18+ years)	10–15 min BI from physician and 6 group psychoeducational sessions (90 min each)
Chang *et al.* (1997) [17]	US general internal medicine clinic	24 women (average age, 39 years)	BI with psychiatrist (duration not reported)
Curry *et al.* (2003) [18]	US urban primary-care clinic	307 men and women (mean, 47 years)	BI from usual physician provider plus and up to 3 telephone follow-up sessions and self-help booklet
Fleming *et al.* (1997) [19]	17 US primary-care practices	774 men and women (18–30 years)	2 15 min BI from usual physician provider 1 month apart plus follow-up phone call from nurse 2 weeks after each visit and self-help booklet
Fleming *et al.* (1999) [20]	24 US primary-care practices	158 men and women (65–75 years)	2 10–15 min BI from usual physician provider 1 month apart plus follow-up phone call from nurse 2 weeks after each visit and self-help booklet
Lock *et al.* (2006) [21]	40 UK primary-care practices	127 men and women (mean, 44 years)	5–10 min BI from nurse and self-help booklet
Maisto *et al.* (2001) [22]	12 US primary-care clinics	301 men and women (mean, 46 years)	Intervention 1: 30–45 min BI from staff with motivational enhancement plus 2 15–20 min follow-up sessions and self-help booklet; Intervention 2: 10–15 min BI from staff and self-help booklet
Nilssen (1991) [23]	University community-health centre in Norway	290 men and 48 women (12–62 years)	Intervention 1: single BI from physician and self-help booklet plus follow-up letter; Intervention 2: more extensive intervention from physician plus self-help booklet and offer of follow-up consultations
Ockene *et al.* (1999) [24]	4 primary-care academic medical centres in the USA	530 men and women (mean, 44 years)	5–10 min BI from trained interventionist and self-help booklet
Richmond *et al.* (1995) [25]	40 primary-care practices in Australia	378 men and women (mean, 38 years)	Intervention 1: 5 min session from physician and self-help booklet plus additional 15–20 min BI 1 week later, additional 5–25 min BI 1 month later and 2 additional 5 min follow-up sessions; Intervention 2: 5 min BI from physician and self-help booklet
Schaus *et al.* (2009) [31]	College student health centre	363 men and women (mean, 20 years)	2 20 min BI sessions from trained providers
Scott & Anderson (1991) [26]	8 primary-care practices in England	226 men and women (mean, 45 years)	10 min BI from physician and self-help booklet
Senft *et al.* (1997) [27]	3 primary-care group practices in the USA	516 men and women (mean, 42 years)	30 s advice from physician, physician's assistant or nurse immediately followed by 15 min BI from health counsellor plus self-help booklet
Wallace *et al.* (1988) [28]	47 primary-care group practices in the UK	909 men and women (17–69 years)	BI from physician plus 1–4 follow-up sessions and self-help booklet
WHO Brief Intervention Study Group (1996) [29]	10 primary-care settings in 8 countries	1260 men (average, 37 years) and 299 women (average, 36 years)	Intervention 1: 5 min BI Intervention 2: 15 min BI with behavioural techniques plus up to 3 follow-up visits

BI, brief intervention.

Table 2. *Alcohol screening and brief intervention trials included in the systematic review that excluded patients with very heavy alcohol use or dependence*

Author (year)	Exclusion criteria
Anderson & Scott (1992) [16]	Heavy use (>105 drinks per week) or advice to cut down in past year
Curry *et al.* (2003) [18]	AUDIT score >15
Fleming *et al.* (1997) [19]	Heavy use (>50 drinks per week), past year alcohol treatment or withdrawal, receipt of physician advice to reduce alcohol use
Fleming *et al.* (1999) [20]	Heavy use (>50 drinks per week), past year alcohol treatment or withdrawal, receipt of physician advice to reduce alcohol use
Lock *et al.* (2006) [21]	AUDIT score >15 for men or >13 for women, or severe dependence
Maisto *et al.* (2001) [22]	Acute dependence symptoms or recent substance abuse treatment
Nilssen (1991) [23]	Dependence
Ockene *et al.* (1999) [24]	In an alcohol treatment program
Richmond *et al.* (1995) [25]	Severe dependence, alcohol-related problems, any past alcohol treatment
Schaus *et al.* (2009) [31]	Consumption of >200 drinks in the last 30 days
Scott & Anderson (1991) [26]	Consumption of >71 drinks per week and/or past year advice to cut down or abstain
Senft *et al.* (1997) [27]	AUDIT score >21
Wallace *et al.* (1988) [28]	Recent health professional advice to reduce drinking and/or gamma-glutamyl-transferase level of 150 IU L^{-1}
WHO Brief Intervention Study Group (1996) [29]	Known or suspected dependence or very high daily consumption, liver damage, prior alcohol treatment, prior health professional warning to abstain

AUDIT, Alcohol Use Disorders Identification Test.

Table 3. *Alcohol screening and brief intervention trials included in the systematic review that did not exclude subjects with very heavy alcohol use or dependence*

Author (year)	Description
Burge *et al.* (1997) [30]	Sample: N = 175 Mexican-American men and women; 65% had abuse and 35% had dependence. Results: no difference in alcohol severity scores by intervention group
Chang *et al.* (1997) [17]	Sample: N = 24 women; 71% had lifetime dependence (58% with current dependence) and 8% had abuse (8% with current abuse); past alcohol treatment was an exclusion. Results: brief intervention not associated with reduced alcohol consumption

dependence found no efficacy for BI in hospitalised patients; 77% of patients in the sample had dependence, because that was the nature of the population found in the hospital [7,40]. A subgroup analysis from the same study found there may have been an impact of BI on drinking among those without dependence only [41]. Another hospital study that excluded patients with dependence or alcohol-related conditions found the effectiveness of BI for decreasing consumption was comparable to that achieved by providing a self-help booklet compared with controls who received no intervention [42]. An additional analysis of BI in the hospital setting by Freyer-Adam *et al.* [43] found no difference in consumption or alcohol-related consequences among the 45% of patients with dependence. Studies that compare BI with more extensive interventions could theoretically be informative, but they have generally not included people identified by screening [6]. In such studies, BI has similar efficacy to longer interventions among those seeking help. It is not known how such results would translate to BI among people with dependence who are not necessarily seeking help.

One might consider the use of categorisation of patients as having dependence a simplification of what is likely a spectrum of severity. Although it is probably true that severity of unhealthy use is on a spectrum rather than a dichotomous phenomenon, it also remains likely that, when patients are categorised as 'dependent', such categorisation includes patients with greater severity. Continuous measures of severity could be used in future studies to better identify those patients for whom BI does or does not have efficacy, but, generally speaking, studies to date have not done so.

This review has methodological strengths as well. Many systematic reviews of BI in primary care have preceded this one, making it likely that few, if any, studies have been missed. Experts in systematic review

methodology have recently encouraged the appropriate inclusion of prior systematic reviews in those that follow [44]. This review focused on the primary-care setting as defined by the US Institute of Medicine and used a wide range of electronic databases. It also excluded studies with substantial methodological limitations.

Based on this review, it is clear that most randomised trials of alcohol BI for screen-identified patients in primary-care settings published to date have excluded patients with very heavy alcohol use or dependence. In the two studies that did include patients with dependence, the efficacy of BI for such patients remains unknown.

Implications

In clinical practice, most do not advocate BI for patients with alcohol dependence, recognising it will likely be insufficient to address this more complex and severe condition. Yet such patients often receive no intervention, whether BI or more extensive treatment. Although BI is expected to lead to referral, treatment initiation, and reductions in alcohol consumption and related problems, these results and those of other studies in other settings suggest this is unlikely. In a study of hospitalised patients by Elvy *et al.* [45], BI decreased alcohol-related consequences and increased treatment enrolment (14% of patients in the BI group enrolled in treatment vs. 4% of patients who got no BI); however, most participants did not have dependence. In a study by Saitz *et al.* involving general hospital patients, no differences in treatment entry were seen, although hypothesis-generating subgroup analyses showed some promise for women and younger men with dependence [40].

Bischof *et al.* [46] compared computerised feedback and telephone-based care with no intervention among primary-care patients with unhealthy alcohol use, including dependence. They found a decrease in heavy drinking only among those with risky use or abuse, but no other outcome differences were observed, and no benefits were observed among patients with dependence. (Specifically, BI did not increase help seeking.) D'onofrio *et al.* [47] conducted a study among emergency department patients without dependence and found no difference in substance abuse or mental health treatment utilisation among those receiving BI versus those receiving no intervention. Finally, although success of drug and alcohol treatment varies depending on rapid availability of treatment, the nature of treatment (e.g. opioid agonist treatment for opioid dependence vs. naltrexone for alcohol dependence) and the availability of support services (e.g. transportation), US programs that provide alcohol screening, BI and referral to treatment (SBIRT) [48] report that most patients with dependence referred to treatment do not accept the referral and, thus, do not enter treatment (Alford D and Smith J, 2009, Personal Communication).

In sum, screening, even when the goal is BI for people with non-dependent unhealthy use, identifies patients with dependence, and the rationale for implementing BI universally among such patients is questionable considering the lack of evidence for its efficacy. We should not 'throw the baby out with the bathwater', however. Clearly, BI has efficacy for primary-care patients with less severe unhealthy alcohol use, and that should continue. The question is whether such screening should be universal if evidence for benefit in an important subgroup is lacking. Some might conclude that it should, because BI will eventually be proven to have efficacy among those with dependence. Others will disagree. Nonetheless, research is needed to determine what, if anything, may have efficacy for patients with alcohol dependence identified by screening in primary care as well as in other health-care settings. Such studies of BI should assess subjects with continuous measures of unhealthy alcohol use severity; be widely generalisable, with few exclusion criteria; and measure important clinical outcomes (e.g. consumption, consequences, cost, referral completion and other health-care utilisation).

The absence of evidence for the efficacy of BI among primary-care patients with screening-identified alcohol dependence raises questions regarding the efficiency of screening and BI, particularly in settings where dependence is common. The finding also highlights the need for developing new approaches to help such patients, particularly if screening and BI are to be disseminated widely.

Acknowledgements

Dr Saitz would like to thank Ms Donna Vaillancourt for editorial preparation of the manuscript for publication. His research is supported in part by support from the US National Institute on Alcohol Abuse and Alcoholism and the National Institute on Drug Abuse (grant numbers R01DA025068 and R01AA10019).

References

[1] Beich A, Thorsen T, Rollnick S. Screening in brief intervention trials targeting excessive drinkers in general practice: systematic review and meta-analysis. BMJ 2003;327:536–42.

[2] Bertholet N, Daeppen JB, Wietlisbach V, Fleming M, Burnand B. Reduction of alcohol consumption by brief alcohol intervention in primary care: systematic review and meta-analysis. Arch Intern Med 2005;165:986–95.

[3] Kaner EF, Dickinson HO, Beyer F, *et al.* The effectiveness of brief alcohol interventions in primary care settings: a systematic review. Drug Alcohol Rev 2009;28:301–23.

[4] Whitlock EP, Polen MR, Green CA, Orleans T, Klein J. Behavioral counseling interventions in primary care to reduce risky/harmful alcohol use by adults: a summary of the evidence for the U.S. Preventive Services Task Force. Ann Intern Med 2004;140:557–68.
[5] Manwell LB, Fleming MF, Johnson K, Barry KL. Tobacco, alcohol, and drug use in a primary care sample: 90-day prevalence and associated factors. J Addict Dis 1998;17:67–81.
[6] Moyer A, Finney JW, Swearingen CE, Vergun P. Brief interventions for alcohol problems: a meta-analytic review of controlled investigations in treatment-seeking and non-treatment-seeking populations. Addiction 2002;97:279–92.
[7] Saitz R, Freedner N, Palfai TP, Horton NJ, Samet JH. The severity of unhealthy alcohol use in hospitalized medical patients. The spectrum is narrow. J Gen Intern Med 2006;21:381–5.
[8] Bischof G, Freyer-Adam J. Brief intervention for medical inpatients with unhealthy alcohol use. Comment. Ann Intern Med 2007;147:589.
[9] US Institute of Medicine Committee on the Future of Primary Care. Primary Care: America's Health in a New Era. Executive Summary. 1996:1. Available at: http://www.nap.edu/nap-cgi/report.cgi?record_id=5152&type=pdfxsum (accessed May 2010).
[10] Harris RP, Helfand M, Woolf SH, *et al*. Current methods of the US Preventive Services Task Force: a review of the process. Am J Prev Med 2001;20:21–35.
[11] Cochrane Drugs and Alcohol Group. Resources for review authors: guidelines to assess study quality, 2003. Available at: http://cdag.cochrane.org/resources-review-authors (accessed May 2010).
[12] Kahan M, Wilson L, Becker L. Effectiveness of physician-based interventions with problem drinkers: a review. CMAJ 1995;152:851–9.
[13] Wilk AI, Jensen NM, Havighurst TC. Meta-analysis of randomized control trials addressing brief interventions in heavy alcohol drinkers. J Gen Intern Med 1997;12:274–83.
[14] Poikolainen K. Effectiveness of brief interventions to reduce alcohol intake in primary health care populations: a meta-analysis. Prev Med 1999;28:503–9.
[15] Ballesteros J, Duffy JC, Querejeta I, Arino J, Gonzalez-Pinto A. Efficacy of brief interventions for hazardous drinkers in primary care: systematic review and meta-analyses. Alcohol Clin Exp Res 2004;28:608–18.
[16] Anderson P, Scott E. The effect of general practitioners' advice to heavy drinking men. Br J Addict 1992;87:891–900.
[17] Chang G, Behr H, Goetz MA, Hiley A, Bigby J. Women and alcohol abuse in primary care. Identification and intervention. Am J Addict 1997;6:183–92.
[18] Curry SJ, Ludman EJ, Grothaus LC, Donovan D, Kim E. A randomized trial of a brief primary-care-based intervention for reducing at-risk drinking practices. Health Psychol 2003;22:156–65.
[19] 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;277:1039–45.
[20] Fleming MF, Manwell LB, Barry KL, Adams W, Stauffacher EA. Brief physician advice for alcohol problems in older adults: a randomized community-based trial. J Fam Pract 1999;48:378–84.
[21] Lock CA, Kaner E, Heather N, *et al*. Effectiveness of nurse-led brief alcohol intervention: a cluster randomized controlled trial. J Adv Nurs 2006;54:426–39.
[22] Maisto SA, Conigliaro J, McNeil M, Kraemer K, Conigliaro RL, Kelley ME. Effects of two types of brief intervention and readiness to change on alcohol use in hazardous drinkers. J Stud Alcohol 2001;62:605–14.
[23] Nilssen O. The Tromso Study: identification of and a controlled intervention on a population of early-stage risk drinkers. Prev Med 1991;20:518–28.
[24] Ockene JK, Adams A, Hurley TG, Wheeler EV, Hebert JR. Brief physician- and nurse practitioner-delivered counseling for high-risk drinkers: does it work? Arch Intern Med 1999;159:2198–205.
[25] Richmond R, Heather N, Wodak A, Kehoe L, Webster I. Controlled evaluation of a general practice-based brief intervention for excessive drinking. Addiction 1995;90:119–32.
[26] Scott E, Anderson P. Randomized controlled trial of general practitioner intervention in women with excessive alcohol consumption. Drug Alcohol Rev 1991;10:313–21.
[27] Senft RA, Polen MR, Freeborn DK, Hollis JF. Brief intervention in a primary care setting for hazardous drinkers. Am J Prev Med 1997;13:464–70.
[28] Wallace P, Cutler S, Haines A. Randomised controlled trial of general practitioner intervention in patients with excessive alcohol consumption. BMJ 1988;297:663–8.
[29] WHO Brief Intervention Study Group. A cross-national trial of brief interventions with heavy drinkers. Am J Public Health 1996;86:948–55.
[30] Burge SK, Amodei N, Elkin B, *et al*. An evaluation of two primary care interventions for alcohol abuse among Mexican-American patients. Addiction 1997;92:1705–16.
[31] Schaus JF, Sole ML, McCoy TP, Mullett N, O'Brien MC. Alcohol screening and brief intervention in a college student health center: a randomized controlled trial. J Stud Alcohol Drugs Suppl 2009;16:131–41.
[32] Altisent R, Cordoba R, Delgado MT, *et al*. Multicenter study on the efficacy of advice for the prevention of alcoholism in primary health care. Med Clin (Barc) 1997;109:121–4.
[33] Huas D, Pessione F, Bouix JC, Demeaux JL, Allemand H, Rueff B. Efficacité à un an d'une intervention brève auprès des consommateurs d'alcool à problèmes. Rev Praticien Méd Gén 2002;16:1343–8.
[34] Fernandez San Martin MI, Bermejo Caja CJ, Alonso Perez M, *et al*. Effectiveness of brief medical counseling to reduce drinkers' alcohol consumption. Aten Primaria 1997;19:127–32.
[35] Diez JF, Pena C, Garcia E, Gaite L. Brief intervention in Cantabria (Spain) in alcohol related problems. [Intervencion breve en Cantabria en problemas relacionados con el alcohol]. Adicciones 2002;14:13–24.
[36] Havard A, Shakeshaft A, Sanson-Fisher R. Systematic review and meta-analyses of strategies targeting alcohol problems in emergency departments: interventions reduce alcohol-related injuries. Addiction 2008;103:368–76.
[37] Nilsen P, Baird J, Mello MJ, *et al*. A systematic review of emergency care brief alcohol interventions for injury patients. J Subst Abuse Treat 2008;35:184–201.
[38] Bernstein E, Bernstein JA, Stein JB, Saitz R. SBIRT in emergency care settings: are we ready to take it to scale? Acad Emerg Med 2009;16:1072–7.
[39] McQueen J, Howe TE, Allan L, *et al*. Brief interventions for heavy alcohol users admitted to general hospital wards. Cochrane Database Syst Rev 2009;3:CD005191.

[40] Saitz R, Palfai TP, Cheng DM, *et al*. Brief intervention for medical inpatients with unhealthy alcohol use: a randomized, controlled trial. Ann Intern Med 2007;146:167–76.
[41] Saitz R, Palfai TP, Cheng DM, *et al*. Some medical inpatients with unhealthy alcohol use may benefit from brief intervention. J Stud Alcohol Drugs 2009;70:426–35.
[42] Holloway AS, Watson HE, Arthur AJ, Starr G, McFadyen AK, McIntosh J. The effect of brief interventions on alcohol consumption among heavy drinkers in a general hospital setting. Addiction 2007;102:1762–70.
[43] Freyer-Adam J, Coder B, Baumeister SE, *et al*. Brief alcohol intervention for general hospital inpatients: a randomized controlled trial. Drug Alcohol Depend 2008;93:233–43.
[44] Whitlock EP, Lin JS, Chou R, Shekelle P, Robinson KA. Using existing systematic reviews in complex systematic reviews. Ann Intern Med 2008;148:776–82.
[45] Elvy GA, Wells JE, Baird KA. Attempted referral as intervention for problem drinking in the general hospital. Br J Addict 1988;83:83–9.
[46] Bischof G, Grothues JM, Reinhardt S, Meyer C, John U, Rumpf HJ. Evaluation of a telephone-based stepped care intervention for alcohol-related disorders: a randomized controlled trial. Drug Alcohol Depend 2008;93:244–51.
[47] D'Onofrio G, Pantalon MV, Degutis LC, *et al*. Brief intervention for hazardous and harmful drinkers in the emergency department. Ann Emerg Med 2008;51:742, 750.e2.
[48] 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;99:280–95.
[49] Aalto M, Saksanen R, Laine P, *et al*. Brief intervention for female heavy drinkers in routine general practice: a 3-year randomized, controlled study. Alcohol Clin Exp Res 2000;24:1680–6.
[50] Aalto M, Seppa K, Mattila P, *et al*. Brief intervention for male heavy drinkers in routine general practice: a three-year randomized controlled study. Alcohol Alcohol 2001;36:224–30.
[51] Cordoba R, Delgado MT, Pico V, *et al*. Effectiveness of brief intervention on non-dependent alcohol drinkers (EBIAL): a Spanish multi-centre study. Fam Pract 1998;15:562–8.
[52] Crawford MJ, Patton R, Touquet R, *et al*. Screening and referral for brief intervention of alcohol-misusing patients in an emergency department: a pragmatic randomised controlled trial. Lancet 2004;364:1334–9.
[53] Fleming M, Brown R, Brown D. The efficacy of a brief alcohol intervention combined with %CDT feedback in patients being treated for type 2 diabetes and/or hypertension. J Stud Alcohol 2004;65:631–7.
[54] Gentilello LM, Rivara FP, Donovan DM, *et al*. Alcohol interventions in a trauma center as a means of reducing the risk of injury recurrence. Ann Surg 1999;230:473–80.
[55] 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;69:243–50.
[56] Heather N, Campion PD, Neville RG, Maccabe D. Evaluation of a controlled drinking minimal intervention for problem drinkers in general practice (the DRAMS scheme). J R Coll Gen Pract 1987;37:358–63.
[57] Israel Y, Hollander O, Sanchez-Craig M, *et al*. Screening for problem drinking and counseling by the primary care physician-nurse team. Alcohol Clin Exp Res 1996;20:1443–50.
[58] Kuchipudi V, Hobein K, Flickinger A, Iber FL. Failure of a 2-hour motivational intervention to alter recurrent drinking behavior in alcoholics with gastrointestinal disease. J Stud Alcohol 1990;51:356–60.
[59] Kunz FM Jr, French MT, Bazargan-Hejazi S. Cost-effectiveness analysis of a brief intervention delivered to problem drinkers presenting at an inner-city hospital emergency department. J Stud Alcohol 2004;65:363–70.
[60] Kypri K, Langley JD, Saunders JB, Cashell-Smith ML, Herbison P. Randomized controlled trial of web-based alcohol screening and brief intervention in primary care. Arch Intern Med 2008;168:530–6.
[61] Lee HS, Mericle AA, Ayalon L, Arean PA. Harm reduction among at-risk elderly drinkers: a site-specific analysis from the multi-site primary care research in substance abuse and mental health for elderly (PRISM-E) study. Int J Geriatr Psychiatry 2009;24:54–60.
[62] Longabaugh R, Woolard RE, Nirenberg TD, *et al*. Evaluating the effects of a brief motivational intervention for injured drinkers in the emergency department. J Stud Alcohol 2001;62:806–16.
[63] Maheswaran R, Beevers M, Beevers DG. Effectiveness of advice to reduce alcohol consumption in hypertensive patients. Hypertension 1992;19:79–84.
[64] Manwell LB, Fleming MF, Mundt MP, Stauffacher EA, Barry KL. Treatment of problem alcohol use in women of childbearing age: results of a brief intervention trial. Alcohol Clin Exp Res 2000;24:1517–24.
[65] McIntosh MC, Leigh G, Baldwin NJ, Marmulak J. Reducing alcohol consumption. Comparing three brief methods in family practice. Can Fam Physician 1997;62:1965–7.
[66] Oslin DW, Grantham S, Coakley E, *et al*. PRISM-E: comparison of integrated care and enhanced specialty referral in managing at-risk alcohol use. Psychiatr Serv 2006;57:954–8.
[67] Pal HR, Yadav D, Mehta S, Mohan I. A comparison of brief intervention versus simple advice for alcohol use disorders in a North India community-based sample followed for 3 months. Alcohol Alcohol 2007;42:328–32.
[68] Persson J, Magnusson PH. Early intervention in patients with excessive consumption of alcohol: a controlled study. Alcohol 1989;6:403–8.
[69] Reinhardt S, Bischof G, Grothues J, John U, Meyer C, Rumpf HJ. Gender differences in the efficacy of brief interventions with a stepped care approach in general practice patients with alcohol-related disorders. Alcohol Alcohol 2008;43:334–40.
[70] Rodriguez-Martos A, Santamarina E, Torralba L, Escayola M, Marti J, Plasencia A. Short-term effectiveness of brief interventions in alcohol-positive traffic casualties. Gac Sanit 2005;19:45–9.
[71] Romelsjo A, Andersson L, Barrner H, *et al*. A randomized study of secondary prevention of early stage problem drinkers in primary health care. Br J Addict 1989;84:1319–27.
[72] Rose HL, Miller PM, Nemeth LS, *et al*. Alcohol screening and brief counseling in a primary care hypertensive population: a quality improvement intervention. Addiction 2008;103:1271–80.
[73] Salaspuro M. Intervention against hazardous alcohol consumption: secondary prevention of alcohol problems. In: Berglund M, Thelander S, Jonsson E, eds. Treating alcohol and drug abuse: an evidence based review. Weinheim: Wiley-VCH, 2003:1–41.

[74] Seppä K. Intervention in alcohol abuse among macrocytic patients in general practice. Scand J Prim Health Care 1992;10:217–22.
[75] Tomson Y, Romelsjo A, Aberg H. Excessive drinking—brief intervention by a primary health care nurse. A randomized controlled trial. Scand J Prim Health Care 1998;16:188–92.

Appendix 1

All searches were limited to English language controlled clinical trials or randomised controlled trials published within a date range of all of 2006 through September 2009.

Databases searched

'New' indicates studies not identified and included in prior systematic reviews.

1. Database of Abstracts of Reviews of Effects (66 studies identified, no new studies included).
2. Cumulative Index to Nursing and Allied Health Literature (21 studies identified, no new studies included).
3. Medline/Cochrane Database of Systematic Reviews (227 studies identified, 8 new studies included).
4. ISI Web of Knowledge (Web of Science) Science Citation Index Expanded/Social Science Citation Index (234 studies identified, no new studies included).
5. PsycINFO (25 studies identified, no new studies included).

Search terms for SCI/SSCI

1. alcohol
2. primary care

Search terms for Medline/Cochrane Database, DARE

1. Settings (combined by 'OR')
   a. family pract$
   b. general pract$
   c. primary care
   d. primary health
   e. family
   f. community
   g. shared care
2. Interventions (combined by 'OR')
   a. brief intervention
   b. alcohol reduction
   c. early intervention
   d. minimal intervention
   e. screening
   f. alcohol therapy
   g. alcohol treatment
   h. harm reduction
   i. counselling
   j. counseling
   k. controlled drinking
   l. brief counselling
   m. brief counseling
   n. physician-based intervention
   o. general practice intervention
   p. secondary prevention
   q. general practitioner's advice
   r. brief physician-delivered counseling
   s. brief nurse-delivered counseling
   t. identification
   u. intervention
3. Topics (combined by 'OR')
   a. alcohol
   b. alcohol (subject heading)
   c. drinking (subject heading)
   d. ethanol (subject heading)
   e. alcohol$ or alcohol consumption

The three categories of search terms were then combined using 'AND'.

Search terms for CINAHL, PsychINFO (combined by 'OR')

1. ('alcohol')
2. (MH 'Alcohol, Ethyl')
3. [MH 'Alcohol Abuse (Saba CCC)']
4. [MH 'Alcohol Abuse Control (Saba CCC)']
5. (MH 'Alcohol Drinking')
6. (MH 'Alcohol-Related Disorders')
7. (MH 'Alcoholic Beverages')
8. (MH 'Alcoholism')
9. (MH 'Alcohol Abuse')
10. (MH 'Alcoholics')
11. [MH 'Risk Control: Alcohol Use (Iowa NOC)']
12. (MH 'Substance Abuse Detection')

The results were then combined with the term 'Primary Care' using 'AND'.

Appendix 2

Studies excluded from the systematic review and reasons for exclusion.

Author (year)	Reason for exclusion
Aalto *et al.* (2000) [49]	Quality: inadequate allocation concealment, >30% of patients lost to follow up, important baseline differences between groups, unclear blinding in outcome assessment
Aalto *et al.* (2001) [50]	Quality: inadequate allocation concealment, >30% of patients lost to follow up, important differences in baseline characteristics between groups, unclear blinding in outcome assessment
Altisent *et al.* (1997) [32]	Language: not published in English
Bischof *et al.* (2008) [46]	Setting: intervention conducted by telephone and via World Wide Web
Cordoba *et al.* (1998) [51]	Quality: 270 of 546 subjects (49%) excluded from analysis because they were lost to follow up or did not adhere to protocol; no intent-to-treat analysis; practice was unit of randomisation, patient was unit of analysis
Crawford *et al.* (2004) [52]	Setting: intervention conducted in an emergency department
Diez *et al.* (2002) [35]	Quality/language: not published in English; randomisation unclear
Fernandez San Martin *et al.* (1997) [34]	Language: not published in English
Fleming *et al.* (2004) [53]	Co-occurring condition: subjects had hypertension or diabetes
Gentilello *et al.* (1999) [54]	Setting: intervention conducted in a trauma centre
Guth *et al.* (2008) [55]	Quality: study lacked a control group
Heather *et al.* (1987) [56]	Quality: post-randomisation exclusions of which the numbers were not reported; less than half of subjects received the full intervention
Huas *et al.* (2002) [33]	Language: not published in English
Israel *et al.* (1996) [57]	Quality: 30% of subjects lost to follow up; baseline comparisons unclear
Kuchipudi *et al.* (1990) [58]	Co-occurring condition: subjects had gastrointestinal disease
Kunz *et al.* (2004) [59]	Setting: intervention conducted in the emergency department
Kypri *et al.* (2008) [60]	Setting: screening and intervention conducted via World Wide Web
Lee *et al.* (2009) [61]	Type: subgroup analysis of Oslin *et al.* (2006) [66]
Longabaugh *et al.* (2001) [62]	Quality/setting: intervention conducted in the emergency department; randomisation unclear; inadequate allocation concealment
Maheswaran *et al.* (1992) [63]	Co-occurring condition: subjects had hypertension
Manwell *et al.* (2000) [64]	Type: subgroup analysis of Fleming *et al.* (1997) [19]
McIntosh *et al.* (1997) [65]	Quality: unclear allocation concealment; baseline characteristic differences between groups; inadequate power
Oslin *et al.* (2006) [66]	Type: compared two interventions, with both groups in active treatment
Pal *et al.* (2007) [67]	Quality: lacked randomisation (alternate group assignment)
Persson & Magnusson (1989) [68]	Quality: consumption measures reported for intervention group but not controls; 31% attrition rate; blinding of participants and outcome assessors unclear
Reinhardt *et al.* (2008) [69]	Type: subgroup analysis by gender of Bischof *et al.* (2008) [46]
Rodriguez-Martos *et al.* (2005) [70]	Language/setting: not published in English; intervention conducted in the emergency department
Romelsjo *et al.* (1989) [71]	Quality: no statistical analysis of results; 151 of 258 patients excluded post randomisation; no intent-to-treat analysis; no adjustment for important differences between groups at baseline
Rose *et al.* (2008) [72]	Quality/co-occurring condition: practice-level intervention lacked randomisation; subjects had hypertension
Salaspuro (2003) [73]	Quality: not peer-reviewed; methodology not adequately reported
Seppä (1992) [74]	Quality/co-occurring condition: randomisation methods unclear; poor follow-up attendance; subjects had macrocytosis
Tomson *et al.* (1998) [75]	Quality: 50%–62% attrition rates; unequal randomisation results (100 in the intervention group, 122 controls); control group not assessed for consumption amounts at baseline; analyses not adjusted for baseline differences

EDITOR'S CHOICE

Editorial

Evidence-based medicine: time for transition and translation (to practice)

Richard Saitz

10.1136/ebm1094

Editor, *Evidence-based Medicine*

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

It is a time of change for evidence-based medicine (EBM, the field) and for *EBM* (the journal). With this issue of *EBM* I become its new Editor; I have assembled an editorial board consisting of primary care and internal medicine practitioners with expertise in both critical appraisal and clinical practice (general practice, family medicine, internal medicine, paediatrics and obstetrics and gynaecology). The main purpose of the journal remains the same: to briefly summarise and critically appraise articles that appear in the peer-reviewed health literature and are likely to be valid and relevant for clinical practice. The world of healthcare, however, has changed since EBM arrived on the scene.

The past 20 years have seen the birth of *EBM*[1] and its adolescence, during which grown-ups in medicine either seemed skeptical about the upstart movement or described it as nothing new. I witnessed this development and took on a role that included its teaching and practice. Journal club during my residency (postgraduate training) usually involved an article that was chosen for little apparent reason (not because of a question it was likely to address). The article was then shredded (at least figuratively) by a resident, and then specialist commentary based on clinical experience and expertise was sprinkled in. When the concept of EBM appeared in the literature, and I had attended a teaching-EBM workshop at McMaster University, my eyes were opened to how much more useful the medical literature could be. The need to handle the information explosion in an organised fashion was great. EBM served as a framework for selecting and evaluating articles. It became a useful tool for keeping up with the literature and a solid basis for practicing medicine.

That science is now mature, EBM firmly in middle age, with experts in searching and critical appraisal all around. And research methodology has advanced, perhaps even outstripping its clinical utility at times. But many challenges remain for EBM on using it to its fullest potential.[2] It is still difficult to keep up with the literature, and for many, critical appraisal is an elusive skill. *EBM* (the journal) helps with these issues. But the real challenges are how to translate evidence into policy and practice. Such translation involves values and preferences. Recent controversies about breast cancer screening and vaccines do not appear to have been as much about evidence as they were about values, preferences, beliefs and translation of evidence into practice and policy (including issues of cost and payment).[3 4] Decisions about what care to pay for vary based on data beyond efficacy, as I recently was reminded during a visit to meet with *British Medical Journal* Group editors in the UK where the news was about how varicella vaccine was not paid for by the National Health Service; in contrast, the vaccine is widely disseminated, considered to be the standard of care and covered by health insurance in the USA. Clinicians and patients (and policymakers) want good evidence, but they also want to know what to do with it. Critical appraisal of an article seems simple in comparison, and practice guidelines help only a little bit in the clinic with individuals.

A few more examples are in order. In this issue of EBM, the reader will find a commentary on the results of a randomised trial of dutasteride for preventing prostate cancer.[5] The study found efficacy for men 50–75 years of age who had had a recent negative prostate biopsy and prostate specific antigen (PSA) level of 2.5 (3.0 if age ≥60) to 10.0 ng/ml. The commentator concludes that the effects of the medication are clinically relevant and that drugs in this class should be considered for men at high risk (eg, like those in the trial). I come to different conclusions from the same evidence. I would emphasise that the effect of dutasteride was limited to low grade cancers, and that we don't know whether prevention of such cancers (often the focus of over-identification and overtreatment) will improve morbidity or reduce mortality. In that context the 5% absolute risk reduction seems of unclear clinical relevance (follow-up biopsies were part of the protocol, not based on symptoms or PSA level). An editorialist wrote that the drug did not prevent prostate cancers; rather it temporarily shrank tumours with low potential for being lethal.[6] He also pointed out that the suppressed PSA levels might delay diagnosis and treatment until the development of high-grade disease. The interpretation of such results depends on much beyond the evidence per se.

I hope international readers will forgive me for a parochial example. I use it because it illustrates the challenges for EBM well. Jacoby Ellsbury is a popular baseball player with the Boston Red Sox. In April this year, another player slammed into him knee first, and a plain radiograph was negative.[7] He was put on the 'disabled list' because of pain in his chest to return when better. Eleven days after the injury he had a CT scan, which he said was done at his request. The CT scan found non-displaced hairline rib fractures, leading a sportswriter to write that "it wasn't just a contusion, as the team discovered yesterday." The team physician pointed out that the test hadn't changed anything – the treatment was the same (rest) – and the player could return when the symptoms subsided. We know a CT scan can identify fractures not seen on plain films. But what is the value of that information? It was of no value to the team physician, but the player (and his fans) felt it was valuable in explaining the duration of symptoms. In fact they seemed surprised that a physician would withhold such testing. Ellsbury said "I'm glad I went about it and did it just to kind of get some closure in what's going on."

What will we do here at *EBM* to help readers keep up and to address the transition of EBM into middle age?

Here is my plan, and I hope to hear from you about it, and whether or not it is meeting your needs.

1. From among the numerous potentially valid peer-reviewed studies published, we will summarise and comment on those most likely to have clinical relevance for medicine practiced by general practitioners, family physicians and internists.
2. We will combine a summary of the original article along with commentary. Original article abstracts can be found online through the link provided. We will summarise the context, key methodological features and results in a structured commentary with subheadings that allow readers to go directly to the section they seek. Experts will provide their view on implications for practice. Our editorial board will peer-review these commentaries.
3. We will include occasional EBM-relevant editorials and perspectives either about articles summarised or about broader issues.
4. *EBM* intends to be a home for 'EBM-ers', those who teach, study and practice EBM. To that end we invite systematic reviews, EBM teaching and research methods articles, 'primer' articles focusing on EBM tools and concepts and an occasional review of what has appeared elsewhere in the literature relevant to EBM (an EBM Roundup).

In addition to this content, I also recognise that, while the paper journal is certainly endearing and has its many fans and users (it will continue), electronic publishing makes it even easier to keep up with the literature through *EBM*. We will make commentaries available online when they are ready, and readers will be able to receive alerts and read them in smaller boluses more frequently if that is their preference.

EBM is at a crossroads, a transition from searching, finding, appraising and keeping up, to translating evidence into policy and practice. It is time to enter the second 20 years during which we will no doubt see EBM fulfil even more of its promise. Hopefully *EBM* will continue to be a useful tool and home for all of you during this next phase. Our plans should help with that as should your comments and recommendations.

References

1. **Guyatt G**. Evidence-based medicine. *ACP J Club (Ann Intern Med)* 1991;**14**(Suppl 2):A-16.
2. **Guyatt G**, Cook D, Haynes B. Evidence based medicine has come a long way. *BMJ* 2004;**329**:990–1.
3. When evidence collides with anecdote, politics, and emotion: breast cancer screening. *Ann Intern Med* 2010;**152**:531–2.
4. Retraction – Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. *Lancet* 2010;**375**:445.
5. **Vickers A**. Dutasteride reduces incident prostate cancer in men after negative prostate biopsy. *Evid based Med*. Published Online First: 2010 Jun 8. [Epub ahead of print].
6. **Walsh PC**. Chemoprevention of prostate cancer. *N Engl J Med* 2010;**362**:1237–8.
7. **Benjamin A**. New tests: Ellsbury Has 4 Fractured Ribs. The Boston Globe. 2010. http://www.boston.com/sports/baseball/redsox/articles/2010/04/23/new_tests_ellsbury_has_4_fractured_ribs/ (Accessed 31 May 2010).

Int J Public Health (2010) 55:527–528
DOI 10.1007/s00038-010-0158-z

LETTER TO THE EDITOR

Most inpatients with unhealthy alcohol use have an alcohol use disorder

Richard Saitz

Published online: 10 June 2010

I read the paper by Bischof et al. (2010) with interest because it addresses an important question with implications for screening and brief intervention. They concluded that at-risk drinkers without alcohol use disorders are "the largest group of unhealthy alcohol consumers" in both hospitals and general practices. They state that our prior study that concluded most inpatients with unhealthy alcohol use had dependence was flawed. I believe they have not interpreted their data and this prior study correctly.

In our hospital study, 81% of those who screened positive had an alcohol disorder (Saitz et al. 2006). Bischof et al. (2010) question these results because (1) they are from one hospital, (2) from a subsample in an intervention study, (3) because of a change in entry criteria during the study, and (4) insurance differences in the US. These are unlikely to explain differences in our findings. In a study in Germany, over half of those who screened positive in the hospital had dependence (after excluding false positives) (Freyer-Adam et al. 2008). In another general hospital study, 66% of those who screened positive in a Barcelona hospital had dependence (Martinez et al. 2007). In our study (Saitz et al. 2006), we were able to screen 99% of 5,813 inpatients who agreed to be screened and 17% were positive. Although compared to those who did not enroll in the clinical trial, those who enrolled (and had diagnostic interviews) were more likely to have Alcohol Use Disorders Identification Test (AUDIT) scores 8 or greater (86 vs. 82%), differences were not statistically significant. Although we did change entry criteria during the study, only one-fifth of subjects were enrolled when the criterion was an AUDIT score of 8 that Bischof et al. (2010) state may have been responsible for discrepant results. In addition, before and after entry criteria changed, the prevalence of unhealthy alcohol use was similar, 19 and 17%, respectively. While lack of insurance was common during this study, our urban safety-net hospital provided care regardless of ability to pay and most admissions are through the emergency department, hospitalizations less sensitive to insurance.

Perhaps more important, the data presented by Bischof et al. (2010) are actually consistent with findings in the literature that suggest that the majority of those who screen positive in hospitals have dependence. In their hospital sample, 15% had unhealthy alcohol use [similar to our and prior studies (Saitz et al. 2006; Roche et al. 2006)], 5.5% had dependence, 2.8% abuse, and 6.2% drank risky amounts; thus, 57% of those with unhealthy use had an alcohol use disorder. This figure is substantially higher than the 33% reported in general practices in their study.

Bischof et al. (2010) are correct that the number of risky drinkers in hospitals is high, and there may be opportunity to intervene, although studies in these settings have been inconclusive, and severity may be part of the explanation, though efficacy has yet to be proven even among those without dependence (Freyer-Adam et al. 2008; McQueen et al. 2009). The literature to date is clear that the majority of patients identified by alcohol screening in hospitals have more severe unhealthy use (an alcohol disorder) than in general practice settings where alcohol brief intervention is known to have efficacy for those without dependence. At a minimum, if screening is implemented in hospitals, clinicians need to be prepared to address substantial numbers of patients they will find with dependence.

R. Saitz (✉)
Boston Medical Center and Boston University,
801 Massachusetts Avenue, Second Floor,
Boston, MA 02118-2393, USA
e-mail: rsaitz@bu.edu

References

Bischof G, Reinhardt S, Freyer-Adam J, Coder B, Grothues JM, Meyer C, Ulrich J, Rumpf HJ (2010) Severity of unhealthy alcohol consumption in medical inpatients and the general population: is the general hospital a suitable place for brief interventions? Int J Public Health. doi:10.1007/s00038-010-0122-y

Freyer-Adam J, Coder B, Baumeister SE, Bischof G, Riedel J, Paatsch K, Wedler B, Rumpf HJ, John U, Hapke U (2008) Brief alcohol intervention for general hospital inpatients: a randomized controlled trial. Drug Alcohol Depend 93(3):233–243

Martinez MB, Rosón B, Hernández R, Lázaro M, Bolao F, Vallejo J, Pujol R (2007) Usefulness of AUDIT-C as screening tool in an opportunistic brief intervention program for alcohol problems in hospitalized patients. http://www.inebria.net/Du14/html/en/dir1338/doc16666.html#w3. Accessed 7 May 2010

McQueen J, Howe TE, Allan L et al (2009) Brief interventions for heavy alcohol users admitted to general hospital wards. Cochrane Database Syst Rev 3:CD005191

Roche AM, Freeman T, Skinner N (2006) From data to evidence, to action: findings from a systematic review of hospital screening studies for high risk alcohol consumption. Drug Alcohol Depend 83:1–14

Saitz R, Freedner N, Palfai TP, Horton NJ, Samet JH (2006) The severity of unhealthy alcohol use in hospitalized medical patients. The spectrum is narrow. J Gen Intern Med 21:381–385

Screening and Brief Intervention for Unhealthy Drug Use in Primary Care Settings: Randomized Clinical Trials Are Needed

Richard Saitz, MD, MPH, Daniel P. Alford, MD, MPH, Judith Bernstein, PhD, ADN, MSN, Debbie M. Cheng, ScD, Jeffrey Samet, MD, MA, MPH, and Tibor Palfai, PhD

Abstract: The efficacy of screening and brief intervention (SBI) for drug use in primary care patients is largely unknown. Because of this lack of evidence, US professional organizations do not recommend it. Yet, a strong theoretical case can be made for drug SBI. Drug use is common and associated with numerous health consequences, patients usually do not seek help for drug abuse and dependence, and SBI has proven efficacy for unhealthy alcohol use. On the other hand, the diversity of drugs of abuse and the high prevalence of abuse and dependence among those who use them raise concerns that drug SBI may have limited or no efficacy. Federal efforts to disseminate SBI for drug use are underway, and reimbursement codes to compensate clinicians for these activities have been developed. However, the discrepancies between science and policy developments underscore the need for evidence-based research regarding the efficacy of SBI for drug use. This article discusses the rationale for drug SBI and existing research on its potential to improve drug-use outcomes and makes the argument that randomized controlled trials to determine its efficacy are urgently needed to bridge the gap between research, policy, and clinical practice.

Key Words: addiction, drug use, primary care, drug screening, brief intervention

(*J Addict Med* 2010;4: 123–130)

Unhealthy drug use is the spectrum from use that risks health consequences (also known as "at-risk use" or "risky use") through dependence. It could be argued that all illicit drug use is unhealthy, because any drug use risks some health or legal consequences. Unhealthy drug use is prevalent in the United States and is associated with numerous health consequences. About 20.4 million Americans (8.3%) aged 12 years and older report past-month illicit drug use, and 2% have a current clinical disorder (ie, abuse or dependence) (Compton et al., 2007; Substance Abuse and Mental Health Services Administration, 2007a). Drug use costs the United States $181 billion per year, primarily because of productivity loss, healthcare costs, and crime (Office of National Drug Control Policy, 2004).

Not all drug use is associated with substance dependence, the most severe disorder. However, for those who develop it, similar to other chronic illnesses, substance dependence is associated with long-term physiologic changes, a relapsing course, variable adherence to care, and the need for ongoing care (McLellan et al., 2000, 2002). In addition to social and legal consequences, co-occurring medical and psychiatric disorders such as depression are common and can trigger relapse (Regier et al., 1990; Kessler et al., 1994; Brindis et al., 1995; Brooner et al., 1997; Ziedonis and Brady, 1997; Friedmann et al., 1998; McLellan et al., 1999; Hasin et al., 2002). Patients with substance dependence are more likely than those without this diagnosis to have myriad conditions including injury, anxiety, psychosis, back pain, headache, arthritis, asthma, peptic disorders, chronic obstructive pulmonary disease, hepatitis C, hypertension, alcoholic gastritis, diseases of the pancreas, and cirrhosis (Mertens et al., 2003; Compton et al., 2007). In addition, the treatment of co-occurring medical and psychiatric conditions in patients with substance dependence is complicated by their risk of poor adherence to medication and other self-care (Golin et al., 2002).

Drug use that does not meet criteria for abuse or dependence can also put people at risk for health consequences. They can not only develop dependence but other medical complications (eg, pneumothorax, myocardial infarction, accidents, and trauma) can also result from such use. Unsafe sex practices and injection drug use are major routes of transmission for human immunodeficiency virus (HIV) (Hudgins et al., 1995; Raj et al., 2007). Mechanisms for increased risk include impaired judgment, increased sex drive, unsafe injection practices, and exchange of sex for drugs. Although risk-reduction interventions in addiction

From the Clinical Addiction Research and Education (CARE) Unit (RS, DPA, DMC, JS), Section of General Internal Medicine, Boston Medical Center and Boston University School of Medicine; Youth Alcohol Prevention Center (RS, TP), Departments of Epidemiology (RS), Social and Behavioral Sciences (JB, JS), Maternal and Child Health (JB), Emergency Medicine (JB), and Biostatistics (DMC), Boston University School of Public Health, Boston, MA; and Department of Psychology (TP), Boston University, Boston, MA.

Received for publication November 18, 2009; accepted March 1, 2010.

Send correspondence and reprint requests to Richard Saitz, MD, MPH, Boston Medical Center, 801 Massachusetts Avenue, 2nd Floor, Boston, MA 02118. e-mail: rsaitz@bu.edu

Supported in part by National Institute on Drug Abuse Grant 1R01DA025068.

ISSN: 1921–0629/10/0403-0123
DOI: 10.1097/ADM.0b013e3181db6b67

treatment settings and sexually transmitted disease clinics have been effective in decreasing these behaviors (Kamb et al., 1998; Lubelczyk et al., 2002; Woody et al., 2003), many people at risk, including those in primary care settings, do not receive such interventions (Wenrich et al., 1997).

However, it is worth noting here that a number of studies have failed to demonstrate health risks associated with drug use in some circumstances. One study found no association between marijuana use and declines in pulmonary function (Tetrault et al., 2007), another found no association between cocaine use and a marker of coronary artery disease (Pletcher et al., 2005), and a third found little evidence of psychological harm associated with drug use among young adults (Macleod et al., 2004). In circumstances in which the risks associated with drug use are small or nonexistent, risk-reduction interventions have had no effect on outcomes.

Clearly, early detection and treatment of drug use that does risk harms could be important if efficacious, yet opportunities for early intervention are limited. To date, the primary focus of treatment has been on persons with more severe unhealthy use; ie, those who meet criteria for substance abuse or dependence. Furthermore, most people with dependence do not seek treatment (Olfson et al., 2000; Compton et al., 2007), and detection and treatment efforts in medical care settings are limited. Thus, reliable methods to screen and treat people who use drugs in the primary care setting have the potential to dramatically improve care and patient outcomes.

SCREENING AND BRIEF INTERVENTION

Screening and brief intervention (SBI) is a comprehensive, integrated, public-health approach to the delivery of early intervention and treatment services for people with the full spectrum of unhealthy substance use. Screening identifies people with unhealthy use and, when followed by an assessment of the severity of substance use, can identify the treatment goal (ie, cutting down or abstinence). Brief intervention (BI) generally involves 1 or 2 counseling sessions of 10 to 30 minutes each, although sessions may be as short as 5 minutes (generally referred to as brief advice) or as long as 1 hour for 4 sessions. Referral may be provided for those identified as needing more extensive specialized treatment.

Primary care centers, hospital emergency departments, and other community health settings see the broadest number and range of patients and thus provide ideal opportunities to screen for and address drug use before more severe consequences occur. Primary care settings provide the best context and opportunities for change over time, because patients have an expectation of preventive care and often have a longitudinal trusting relationship with a clinician.

Conceptual Framework

Although BI includes clear directive advice, focus is primarily on increasing patient insight and awareness regarding substance use and encouraging behavioral change through motivational interviewing (MI) and self-management approaches (Miller, 1983; Miller and Rollnick, 1991; Heather, 1995). MI is based on psychological theories of attitude and behavior change (Rogers, 1957; Becker, 1974; Bandura, 1977; Miller, 1983; Prochaska and DiClemente, 1986; Miller and Rollnick, 1991), addressing the fact that patients frequently neither recognize their health behaviors as hazardous nor acknowledge a desire to change (Rollnick et al., 1992). Factors that enhance willingness and ability for behavior change have roots in self-management (Mahoney, 1979), self-control (Carver and Scheier, 1982), and self-regulation (Kanfer, 1986) theories that describe how individuals plan, guide, and monitor behavior. A number of these factors have been used successfully in interventions for unhealthy alcohol use, such as altering norms and standards (Dimeff et al., 1999) specifying change plans (Sobell and Sobell, 1983), and increasing the probability of action by helping the patient take the first step, such as facilitating referral to treatment (Heather, 1995).

Elements of effective BIs include Feedback on personal risk, emphasis on Responsibility, clear Advice, a Menu of change options, clinician Empathy, and facilitation of patient Self-efficacy (FRAMES) (Bien et al., 1993). BI models tested in primary care have been delivered by physicians, nurses, health educators, advocates, computer, or pamphlet. Each involved feedback, advice, goal setting, and follow-up. Models may differ by how, by whom, and in what context they are delivered; therefore, different training, supervision, and quality monitoring are required (Miller et al., 2005). Although patient and interventionist characteristics are important, feasibility and cost are particularly relevant to effective implementation in (usually) busy general health settings with numerous other priorities.

SBI for Unhealthy Alcohol Use

SBI for unhealthy alcohol use has been described in the scientific literature for nearly 50 years. In 1961, Chafetz (1961) found that subjects with alcoholism in an emergency department were significantly more likely to follow-up in an alcohol clinic after brief advice from a psychiatrist than after no advice (42% vs 1%). Later studies (Edwards et al., 1977, 1983) found that BI was as efficacious as more intensive treatment (although this finding was likely attributable to studying treatment-seeking patients rather than those identified by screening—a critical distinction). Thirty years after Chafetz published his findings, Bien et al. (1993) reviewed 32 studies that showed BI effectively reduced unhealthy alcohol use, and meta-analyses have confirmed its efficacy for non-dependent unhealthy alcohol use in primary care settings leading to a universal screening practice guideline in the United States (Wilk et al., 1997; Moyer et al., 2002; Beich et al., 2003; Ballesteros et al., 2004; US Preventive Services Task Force, 2004; Bertholet et al., 2005; Kaner et al., 2009).

Such evidence-based guidelines, as is appropriate for universal preventive services, only appeared after randomized trials provided supportive evidence. These trials most often involved primary care clinicians delivering BIs, although in some cases, interventions were conducted by other healthcare professionals hired specifically to deliver them. Some of the most notable studies found that more than 1 contact improved efficacy (Fleming et al., 1997; Longabaugh et al., 2001; McKay et al., 2005; Brown et al., 2007).

Despite this relatively robust evidence, several studies show that BI was not effective in hospitalized patients (Em-

men et al., 2004; McQueen et al., 2009), in largely alcohol-dependent patients with prevalent use of other drugs (Saitz et al., 2007), in emergency departments (Daeppen et al., 2007; D'Onofrio et al., 2008; Havard et al., 2008; Nilsen et al., 2008), and in some general practice settings (Richmond et al., 1995; Beich et al., 2007). Factors such as sex, age, homelessness, and cognitive status influence effectiveness (Saitz et al., 2006), and SBI has not been effective in linking medical inpatients with treatment for alcohol dependence after discharge (Saitz et al., 2007). The best evidence for alcohol BI is for reductions in consumption (in contradistinction to consequences) among patients in the primary care setting who have unhealthy use that is not severe (Kaner et al., 2009).

SBI for Unhealthy Drug Use

Although randomized controlled trials have proven the efficacy of SBI for nondependent unhealthy alcohol use in primary care settings (US Preventive Services Task Force, 2004; Kaner et al., 2009), the evidence is much more limited regarding its effectiveness for other drug use. Although the prevalence of drug use in primary care is variable, it is much lower than that of unhealthy alcohol use. Estimates range from 3% of adults reporting past-year use in a Health Maintenance Organization setting (Mertens et al., 2005), to 5% reporting past 90-day use in practices in Wisconsin (Manwell et al., 1998), to 8% reporting past-year use in an urban practice (among whom only 22% received treatment) (Olfson et al., 2000). Marijuana use is especially common: in 1 study, 8% of young women used marijuana monthly (Rose et al., 2007); in another, 9% to 17% of adults report past 6-month use (Pasternak and Fleming, 1999).

Because most US adults (83%) report having an outpatient visit in the past year (Pleis and Lethbridge-Çejku, 2007), primary care settings provide a natural setting to pursue health behavior change, including unhealthy drug use. Because of this, US policymakers have sought to make SBI, referral, and treatment (Substance Abuse and Mental Health Services Administration, 2007b) an important part of addressing the nation's drug problems, and reimbursement codes for insurers to compensate clinicians for these activities have been approved by the American Medical Association (Anonymous, 2006). However, scarcity of evidence from controlled clinical trials in the primary care setting (or in any setting among those identified by screening in contradistinction to those seeking help) has prevented the inclusion of drug SBI in preventive service recommendations (US Preventive Services Task Force, 2008). No major professional organizations recommend universal drug SBI in primary care settings, and its dissemination has been limited mainly to externally funded programs that specifically support the activity. To date, few randomized trials have addressed the question of whether SBI reduces illicit drug use and its consequences when identifying patients who need treatment before they seek it. They have also not adequately explored whether the benefits of SBI outweigh potential harms such as increased use, consequences of breached confidentiality, or stigma.

Screening Tools for Unhealthy Drug Use

One reason for the scarcity of SBI research in primary care may be the lack (perhaps until recently) of brief and valid screening instruments for substances other than alcohol or tobacco. Screening tests have been used for unhealthy drug use, however, most have been alcohol tests modified for drug use, have focused on dependence, or have not been validated extensively. For example, the Drug Abuse Screening Test (DAST) detects drug use problems and has been widely used in federal SBI programs, but it was not validated in primary care until recently (Skinner, 1982; Smith et al., 2010).

In recent years, screening instruments for drug use have been developed and validated in primary care settings. The Alcohol Smoking and Substance Involvement Screening Test (ASSIST) (Humeniuk et al., 2008a) is used for tobacco, alcohol, cannabis, cocaine, stimulants, sedatives, hallucinogens, opioids, and several other drugs. Although it does not directly identify risky amounts of alcohol consumption (a substantial clinical limitation), it has excellent concurrent validity compared with the Addiction Severity Index-Lite version, the Severity of Dependence Scale, and the Alcohol Use Disorders Identification Test (Humeniuk et al., 2008b) as well as construct validity, test-retest reliability, discrimination of severity, and sensitivity and specificity for a drug use disorder and, in some cases, any drug use (cocaine, amphetamine, benzodiazepines, and opioids) (Newcombe et al., 2005; Humeniuk et al., 2008b). In a Brazilian study of the ASSIST that included 99 patients from mostly primary care settings and 48 patients in drug treatment, sensitivity and specificity were 84% and 98%, respectively, for a cocaine use disorder, and 87% and 95%, respectively, for a marijuana disorder (Henrique et al., 2004). However, an important drawback of the ASSIST is that it may be too lengthy to be feasible in busy primary care settings.

In addition to the ASSIST, Smith et al. (2010) evaluated a single-item screening tool—"How many times in the past year have you used an illegal drug or used a prescription medication for nonmedical reasons?"—among 286 primary care patients in a large urban hospital-based setting. Thirty-five percent screened positive for any drug use, and 13% (more than one-third of those who screened positive) met criteria for a current drug use disorder. A response of "1 or more" was 100% sensitive and 74% specific for a drug use disorder, 94% sensitive and 91% specific for use with consequences, and 93% sensitive and 94% specific for any drug use. This single-item tool has promise, although it has been validated in only 1 primary care practice. Other brief tools with limited validation have generally focused on disorders and not the spectrum of unhealthy use (Brown and Rounds, 1995; Brown et al., 2001) and often combine both alcohol and drugs. The availability of brief validated screening tools is an important foundation for increased research on the effectiveness of SBI for drug use in primary care.

EFFICACY OF DRUG SBI: CURRENT EVIDENCE

Substantial evidence in nonprimary healthcare settings and under different circumstances (eg, among people actively seeking help) informs the question of whether drug SBI has

efficacy in primary care but does not establish definitive answers (Bashir et al., 1994; Cormack et al., 1994; Stephens et al., 2000; Copeland et al., 2001; Babor, 2004; McCambridge and Strang, 2004, 2005; Baker et al., 2005; Denis et al., 2006; Carroll et al., 2006; Voshaar et al., 2006; Ball et al., 2007). To our knowledge, no randomized controlled trials of drug SBI in adult primary care settings have been published in the peer-reviewed literature.

Madras et al. (2009) conducted a before/after retrospective uncontrolled study to evaluate the 6-state SBI referral and treatment initiative of the Center for Substance Abuse Treatment. Settings were diverse, including trauma centers, emergency departments, primary and specialty care sites, and hospitals. Of the 459,599 patients screened for the study, 23% tested positive for risky or problematic alcohol or drug use. Of these, 70% had screening results that suggested BI would be a reasonable course of action, 14% were recommended for brief treatment, and 16% had screening results that suggested they should be referred to specialty substance dependence treatment. How many patients actually received intervention or treatment is unknown. Ten percent of patients who screened positive were randomly selected for reassessment 6 months later, at which time self-reported rates of heavy alcohol use and illicit drug use had decreased by 39% and 68%, respectively. Self-reported rates of overall health, employment, housing status, and criminality among persons who were in categories in which they should have been offered brief treatment or referral had also improved significantly.

In a landmark randomized controlled study of BI in adult outpatients with cocaine or heroin use identified by screening, Bernstein et al. (2005) screened 23,660 patients from women's health, homeless, and urgent care clinics and randomized those who screened positive for risky cocaine or heroin use (N = 1175) to a brief negotiated interview (BNI) or to receipt of a referral list and written advice. Although a homeless clinic and women's health clinics could be considered primary care settings, urgent care settings are different from primary care, and subgroup analyses by site are not available. Ninety-five percent of eligible subjects were enrolled in the study, and 82% were available for follow-up. Post hoc, 19% of those followed up were excluded because baseline drug use was not confirmed biochemically. At 6 months, abstinence was documented among 40% of the intervention subjects and 31% of control subjects who used opiates at baseline, and 22% of the intervention subjects and 17% of the control subjects who used cocaine at baseline (statistically significant differences). No difference in receipt of help (90% of which was detoxification) was observed between groups.

In an uncontrolled study by Bernstein et al. (1997), patients who screened positive for substance problems in the emergency department were given a BNI. At 60 to 90 day follow-up (completed by 8% of those who screened positive for alcohol or drugs), patients who received the BNI had significant reductions in substance use, including a 45% reduction in drug-problem severity. The number of referrals and receipt of addiction treatment also quadrupled from 6% to 23% after a BNI.

The Health Evaluation and Linkage to Primary Care study linked 470 drug and alcohol abusers at a detoxification unit to primary medical care and assessed the effect during a 2-year period, during which 85% completed at least 1 follow-up assessment (Samet et al., 1996, 2003). Results showed that a brief multidisciplinary intervention could link people with primary medical care and that primary care exposure was associated with greater drug abstinence (Saitz et al., 2005). What component led to improvement, and whether BI was a factor, is not known.

Finally, a number of studies have suggested that BI may decrease substance use among teens. In a randomized trial of adolescents with recent drug use (N = 59) in a primary care setting in Brazil, BI decreased ecstasy and marijuana use and related drug problems (De Micheli et al., 2004). In the United States, Project CHAT examined the effect of a brief MI intervention on alcohol consumption and drug use for high-risk teens in a primary care clinic (D'Amico et al., 2008). Teens who screened positive for negative consequences related to substance use were randomized to receive either a brief MI intervention or to a control group (care as usual). Participants in the intervention group reported less marijuana use, lower perceived prevalence of marijuana use, fewer friends who used marijuana, and decreased intent to use marijuana in the next 6 months compared with controls.

In 2 randomized controlled trials by Tait et al. (2004, 2005), BI among adolescents in an emergency department increased drug-treatment attendance and reduced return visits for consequences related to substance use. In another study, a single MI-style feedback session decreased some drug use (but not alcohol or marijuana use) among homeless adolescents (Peterson et al., 2006), and additional studies have shown its efficacy for youth in mandated treatment and in high schools (White et al., 2006; Srisurapanont et al., 2007; Winters and Leitten, 2007). In a pediatric emergency department, BI for marijuana use resulted in greater levels of abstinence at 1 year compared with controls who received only written advice (Bernstein et al., 2009). A randomized trial in a hospital assessed effects of 2 counseling sessions on psychoactive prescription drug use and found that intervention was associated with decreased use. However, some of the subjects had regular use but not abuse, and it is not clear whether the decrease was beneficial because some patients were taking pain medication regularly for pain (Zahradnik et al., 2009).

Reasons for Caution

Although SBI has proven efficacy for nondependent unhealthy alcohol and drug use in some healthcare settings and populations, this benefit may not translate to drug users identified by screening in primary care (or elsewhere) due to a number of clinical concerns and challenges. In a general health setting, BI for drug use is likely to be more complicated than BI for alcohol use and is likely to involve a greater proportion of patients with dependence than is BI for screen-identified unhealthy alcohol use.

The many patients who use more than 1 drug or use alcohol and another drug make BI more complicated than it is for alcohol alone (Falk et al., 2006; McCabe et al., 2008). These drugs have variable forms, costs, risks, consequences, and ways for clinicians to identify use. For example, in our experience implementing drug SBI clinically, a BI for depen-

dent injection heroin use, with its attendant risks of overdose and HIV infection, is different from a BI for occasional users of marijuana who perceive their use to be without risk or even beneficial to their health. Most abused drugs are illegal or used illegally, which can complicate addressing their use in medical settings by raising patient and physician concerns about confidentiality and medical record documentation. Prescription drug abuse presents additional challenges as clinicians struggle to distinguish between appropriate and inappropriate use.

Another clinical concern is that a larger proportion of patients with drug use identified by screening will have dependence compared with those identified as having unhealthy alcohol use (Substance Abuse and Mental Health Services Administration, 2007b). BI, even with a goal of referral, has not been proven even for alcohol in such circumstances, and is not widely recommended as the sole intervention for dependence (Moyer et al., 2002; McQueen et al., 2009).

In addition to the aforementioned clinical concerns, the state of the evidence regarding screening is an additional reason for caution. Limited availability of feasible brief screening tools presents a significant barrier to implementation. Few validation studies have been conducted in general health settings on the screening tools discussed herein, and the DAST and ASSIST cannot be considered brief, having 10 to 80 or more items. Given the known challenges to implement SBI for alcohol with 1- to 3-item screening tests, the DAST and ASSIST are not likely to be disseminated widely, even if extensively validated. Although the single-item screening tool discussed earlier has the potential to minimize this barrier to implementing drug BI, further validation is needed.

With regard to BI, the lack of randomized trial evidence along with the inability to generalize results among treatment seekers compared with those identified by screening are additional concerns surrounding universal drug SBI. The assumption that drug-treatment efficacy among those seeking help will apply to people identified by screening and not necessarily seeking treatment is likely inaccurate. Assessing whether BI has efficacy among people identified by screening—the common situation in primary care—is important, because these patients present with varying levels of readiness to change and a range of drug-use severity. Unfortunately, this distinction is challenging to test empirically because randomizing these 2 patient populations to the same treatment and control groups and comparing the effects would prove difficult, and people with less severe unhealthy use are not likely to be well represented in treatment-seeking populations. For a condition such as drug use in which motivation plays an important role, it seems logical that BI (ie, counseling that addresses motivation to change) would have different outcomes among those seeking help versus those not seeking it.

Aside from the challenge of translating study results from patients seeking help to those identified by screening, the absence of randomized trial evidence for drug SBI among adults in the primary care setting is a major concern. Observational studies and uncontrolled trials have limited ability to establish causality and thus cannot provide sufficient evidence to support recommendations for universal implementation of drug SBI. Improvements in the range of 40% to 70% seen in such studies (Bernstein et al., 1997; Madras et al., 2009) may be the result of many factors besides BI, including regression to the mean, natural history of drug use after a patient-initiated voluntary healthcare contact, and confounding by prognostic factors that change across time. Effects seen in randomized trials are much more modest (about one-tenth of the magnitude).

In the only randomized controlled trial of SBI in a primary care setting (aside from the Brazilian study involving adolescents described earlier [De Micheli et al., 2004]), World Health Organization researchers who developed the ASSIST conducted a 5-country Phase III randomized trial of BI among 731 persons who screened positive on the ASSIST for risky cannabis, cocaine, amphetamine, or opioid use. The results appear in a technical report (Humeniuk et al., 2008a) not yet published in a peer-reviewed journal. Patients recruited from sexually transmitted disease clinics, walk-in clinics, a dental clinic, and community medical care sites (only some of which could be considered primary care) were randomly assigned to either BI or no counseling. Although BI reduced substance-use scores in a preliminary single-site subgroup in the study (Newcombe et al., 2005), results of the larger trial were less conclusive (Humeniuk et al., 2008a). Differences between the 2 groups were small and of unknown clinical importance (3 points on a scale with a maximum value of 336 points), effects were seen for cannabis and stimulants but not for opioids, and substance use was not significantly impacted at the US site. Although the authors speculate that the lack of efficacy in the US was due to informed-consent procedures having an intervention effect, numerous alcohol SBI studies in the United States that had informed-consent procedures have found SBI effects (Whitlock et al., 2004). In summary, intervention effects in this study were not convincingly significant, nor were most patients recruited from settings that could be considered primary care (ie, longitudinal, continuous, comprehensive care). As such, the study does not settle the question of whether SBI for drug use is of clinical benefit in primary care.

In addition, effects seen in trials that involve substantial training and effort to maintain intervention fidelity may not translate into real-world clinical practice. Effects in practice are likely to be smaller than those seen in research studies, which are small to begin with. BI for alcohol, a less complex clinical problem than drug use, among those identified by screening is associated with a 10% to 12% absolute decrease in risky use (Whitlock et al., 2004; Bertholet et al., 2005; Kaner et al., 2009). In the randomized trial by Bernstein et al. (2005) among outpatients who screened positive for cocaine or heroin use, BI was associated with a 5% to 9% increase in abstinence. These small effects could be wiped out in practice if training and fidelity maintenance are not as good as in controlled trials in which clinicians or other interventionists are trained by study personnel for study purposes. The issue of translating efficacious interventions into practice also raises feasibility concerns if the commitment of clinical staff to delivering BI is uneven or inadequate.

Conversely, results in controlled research settings could be smaller than those observed in clinical practice because of assessment effects (ie, the notion that research assessments alone may lead to behavioral change much as an intervention might) among control group patients. However, assessment effects are unlikely to explain the large improvements shown in some nonrandomized studies for such a recalcitrant clinical condition as a drug use disorder. In fact, no assessment effects were found in at least some BI studies that tested for them (Daeppen et al., 2007).

CONCLUSION

Although SBI can occur in many settings and can target alcohol, drugs, or both, determining its efficacy and feasibility in primary care requires rigorous testing of brief screening tools and of different models of SBI. This issue, at a key clinical and policy crossroads, is of great importance given the severity and cost of the drug problem in the United States. The discrepancy between policy developments (reimbursement codes for drug SBI and a large federal SBI grant-funded program that includes drugs) and the existing evidence base for drug SBI underscore the need for randomized clinical trials to determine its effectiveness in primary care. Taking into consideration the perspectives of national professional societies, quality measurement groups, and practice guideline developers (none of which has come out in support of drug SBI), current policy and practice—at least as part of federally funded SBI programs—have gone well beyond the evidence base. Existing studies are insufficient to justify changes in clinical practice, just as decades of alcohol SBI data did not move practice guidelines in the United States until the completion of 2 large randomized controlled trials (Fleming et al., 1997; Ockene et al., 1999; US Preventive Services Task Force, 2004). Validated screening tools exist and can be further tested, and most adults visit primary medical care settings in which drug problems can be identified and BI can be conducted.

Although randomized clinical trials are challenging to implement, particularly with drug-using persons and those of lower socioeconomic status, including homelessness, high follow-up rates can be achieved (Samet et al., 1996, 2003; Saitz et al., 2003, 2005; Bernstein et al., 2005). Findings from a pragmatic trial in primary care, enrolling people who use drugs and who are at risk for or have experienced related consequences, are necessary to determine whether BI should be widely disseminated to reduce the national burden of drug-related illness and other negative effects, including the spread of HIV. Efficacious BI models with favorable economic characteristics have the potential to significantly reduce the national burden of drug use and consequences. Conversely, lack of efficacy or excessive cost would force reconsideration of drug SBI as a broadly applicable strategy and would, appropriately, redirect efforts to address the problem.

Given that even proven strategies of SBI (eg, SBI for alcohol in primary care settings) have not been widely implemented in practice, efficacy studies should include elements that can inform real-world effectiveness and implementation. Features might include minimizing restrictive entry criteria, recruiting subjects from diverse populations, minimizing intensity of study procedures to improve retention, and testing strategies that can be reproduced and financed in clinical practice settings. It is time for US efficacy studies of drug SBI in primary care settings that test models feasible in the real world and consider costs, sustainability, and outcomes.

ACKNOWLEDGMENTS

The authors thank Donna Vaillancourt for editorial preparation of the manuscript for publication. This article was supported in part by Grant Number 1R01DA025068 from the US National Institute on Drug Abuse (NIDA).

REFERENCES

Anonymous. New codes could encourage more screening and brief intervention. *Alcohol Drug Abuse Wkly* 2006;18:6.

Babor T. Brief treatments for cannabis dependence: Findings from a randomized multisite trial. *J Consult Clin Psychol* 2004;72:455–466.

Baker A, Lee NK, Claire M, et al. Brief cognitive behavioural interventions for regular amphetamine users: A step in the right direction. *Addiction* 2005;100:367–378.

Ball SA, Martino S, Nich C, et al. Site matters: Multisite randomized trial of motivational enhancement therapy in community drug abuse clinics. *J Consult Clin Psychol* 2007;75:556–567.

Ballesteros J, Gonzalez-Pinto A, Querejeta I, et al. Brief interventions for hazardous drinkers delivered in primary care are equally effective in men and women. *Addiction* 2004;99:103–108.

Bandura A. Self-efficacy: Toward a unifying theory of behavioral change. *Psychol Rev* 1977;84:191–215.

Bashir K, King M, Ashworth M. Controlled evaluation of brief intervention by general practitioners to reduce chronic use of benzodiazepines. *Br J Gen Pract* 1994;44:408–412.

Becker MH. The Health Belief Model and Personal Health Behavior. Thorofare, NJ: Charles B. Slack; 1974.

Beich A, Gannik D, Saelan H, et al. Screening and brief intervention targeting risky drinkers in Danish general practice—A pragmatic controlled trial. *Alcohol Alcohol* 2007;42:593–603.

Beich A, Thorsen T, Rollnick S. Screening in brief intervention trials targeting excessive drinkers in general practice: Systematic review and meta-analysis. *BMJ* 2003;327:536–542.

Bernstein E, Bernstein J, Levenson S. Project ASSERT: An ED-based intervention to increase access to primary care, preventive services, and the substance abuse treatment system. *Ann Emerg Med* 1997;30:181–189.

Bernstein E, Edwards E, Dorfman D, et al. Screening and brief intervention to reduce marijuana use among youth and young adults in a pediatric emergency department. *Acad Emerg Med* 2009;16:1174–1185.

Bernstein J, Bernstein E, Tassiopoulos K, et al. Brief motivational intervention at a clinic visit reduces cocaine and heroin use. *Drug Alcohol Depend* 2005;77:49–59.

Bertholet N, Daeppen JB, Wietlisbach V, et al. Reduction of alcohol consumption by brief alcohol intervention in primary care: Systematic review and meta-analysis. *Arch Intern Med* 2005;165:986–995.

Bien TH, Miller WR, Tonigan JS. Brief interventions for alcohol problems: A review. *Addiction* 1993;88:315–335.

Brindis C, Pfeffer R, Wolfe A. A case management program for chemically dependent clients with multiple needs. *J Case Manag* 1995;4:22–28.

Brooner RK, King VL, Kidorf M, et al. Psychiatric and substance use comorbidity among treatment-seeking opioid abusers. *Arch Gen Psychiatry* 1997;54:71–80.

Brown RL, Leonard T, Saunders LA, et al. A two-item conjoint screen for alcohol and other drug problems. *J Am Board Fam Pract* 2001;14:95–106.

Brown RL, Rounds LA. Conjoint screening questionnaires for alcohol and other drug abuse: Criterion validity in a primary care practice. *Wis Med J* 1995;94:135–140.

Brown RL, Saunders LA, Bobula JA, et al. Randomized-controlled trial of a telephone and mail intervention for alcohol use disorders: Three-month drinking outcomes. *Alcohol Clin Exp Res* 2007;31:1372–1379.

Carroll KM, Ball SA, Nich C, et al. Motivational interviewing to improve treatment engagement and outcome in individuals seeking treatment for

substance abuse: A multisite effectiveness study. *Drug Alcohol Depend* 2006;81:301–312.
Carver CS, Scheier MF. Control theory: A useful conceptual framework for personality-social, clinical, and health psychology. *Psychol Bull* 1982; 92:111–135.
Chafetz ME. A procedure for establishing therapeutic contact with the alcoholic. *Q J Stud Alcohol* 1961;22:325–328.
Compton WM, Thomas YF, Stinson FS, et al. Prevalence, correlates, disability, and comorbidity of DSM-IV drug abuse and dependence in the United States: Results from the national epidemiologic survey on alcohol and related conditions. *Arch Gen Psychiatry* 2007;64:566–576.
Copeland J, Swift W, Roffman R, et al. A randomized controlled trial of brief cognitive-behavioral interventions for cannabis use disorder. *J Subst Abuse Treat* 2001;21:55–64.
Cormack MA, Sweeney KG, Hughes-Jones H, et al. Evaluation of an easy, cost-effective strategy for cutting benzodiazepine use in general practice. *Br J Gen Pract* 1994;44:5–8.
D'Amico EJ, Miles JN, Stern SA, et al. Brief motivational interviewing for teens at risk of substance use consequences: A randomized pilot study in a primary care clinic. *J Subst Abuse Treat* 2008;35:53–61.
D'Onofrio G, Pantalon MV, Degutis LC, et al. Brief intervention for hazardous and harmful drinkers in the emergency department. *Ann Emerg Med* 2008;51:742–750.
Daeppen JB, Gaume J, Bady P, et al. Brief alcohol intervention and alcohol assessment do not influence alcohol use in injured patients treated in the emergency department: A randomized controlled clinical trial. *Addiction* 2007;102:1224–1233.
De Micheli D, Fisberg M, Formigoni ML. [Study on the effectiveness of brief intervention for alcohol and other drug use directed to adolescents in a primary health care unit]. *Rev Assoc Med Bras* 2004;50:305–313.
Denis C, Lavie E, Fatseas M, et al. Psychotherapeutic interventions for cannabis abuse and/or dependence in outpatient settings. *Cochrane Database Syst Rev* 2006(3):CD005336.
Dimeff LA, Baer JS, Kivlahan DR, et al. Brief Alcohol Screening and Intervention for College Students: A Harm Reduction Approach. New York: Guilford Press; 1999.
Edwards G, Duckitt A, Oppenheimer E, et al. What happens to alcoholics? *Lancet* 1983;2:269–271.
Edwards G, Orford J, Egert S, et al. Alcoholism: A controlled trial of "treatment" and "advice." *J Stud Alcohol* 1977;38:1004–1031.
Emmen MJ, Schippers GM, Bleijenberg G, et al. Effectiveness of opportunistic brief interventions for problem drinking in a general hospital setting: Systematic review. *BMJ* 2004;328:318.
Falk DE, Yi HY, Hiller-Sturmhöfel S. An epidemiologic analysis of co-occurring alcohol and tobacco use and disorders: Findings from the National Epidemiologic Survey on Alcohol and Related Conditions. *Alcohol Res Health* 2006;29:162–171.
Fleming MF, Barry KL, Manwell LB, et al. Brief physician advice for problem alcohol drinkers: A randomized controlled trial in community-based primary care practices. *JAMA* 1997;277:1039–1045.
Friedmann PD, Saitz R, Samet JH. Management of adults recovering from alcohol or other drug problems: Relapse prevention in primary care. *JAMA* 1998;279:1227–1231.
Golin CE, Liu H, Hays RD, et al. A prospective study of predictors of adherence to combination antiretroviral medication. *J Gen Intern Med* 2002;17:756–765.
Hasin D, Liu X, Nunes E, et al. Effects of major depression on remission and relapse of substance dependence. *Arch Gen Psychiatry* 2002;59:375–380.
Havard A, Shakeshaft A, Sanson-Fisher R. Systematic review and meta-analyses of strategies targeting alcohol problems in emergency departments: Interventions reduce alcohol-related injuries. *Addiction* 2008; 103:368–376.
Heather N. Brief intervention strategies. In: Hester RK, Miller WR, eds. Handbook of Alcoholism Treatment Approaches: Effective Alternatives. Needham Heights, MA: Allyn and Bacon; 1995.
Henrique IF, De Micheli D, Lacerda RB, et al. [Validation of the Brazilian version of Alcohol, Smoking and Substance Involvement Screening Test (ASSIST)]. *Rev Assoc Med Bras* 2004;50:199–206.
Hudgins R, McCusker J, Stoddard A. Cocaine use and risky injection and sexual behaviors. *Drug Alcohol Depend* 1995;37:7–14.
Humeniuk R, Ali R, Babor TF, et al. Validation of the Alcohol, Smoking And Substance Involvement Screening Test (ASSIST). *Addiction* 2008a;103:1039–1047.
Humeniuk R, Dennington V, Ali R. The Effectiveness of a Brief Intervention for Illicit Drugs Linked to the Alcohol, Smoking and Substance Involvement Screening Test (ASSIST) in Primary Health Care Settings: A Technical Report of Phase III Findings of the WHO ASSIST Randomized Controlled Trial. Geneva, Switzerland: World Health Organization, 2008b. Available at: http://www.who.int/substance_abuse/activities/assist_technicalreport_phase3_final.pdf. Accessed January 26, 2010.
Kamb ML, Fishbein M, Douglas JM Jr, et al. Efficacy of risk-reduction counseling to prevent human immunodeficiency virus and sexually transmitted diseases: A randomized controlled trial. Project RESPECT Study Group. *JAMA* 1998;280:1161–1167.
Kaner EF, Dickinson HO, Beyer F, et al. The effectiveness of brief alcohol interventions in primary care settings: A systematic review. *Drug Alcohol Rev* 2009;28:301–323.
Kanfer F. Implications of a self-regulation model of therapy for treatment of addictive behaviors. In: Miller W, Heather N, eds. Treating Addictive Behaviors: Processes of Change. New York: Plenum Press; 1986.
Kessler RC, McGonagle KA, Zhao S, et al. Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. Results from the National Comorbidity Survey. *Arch Gen Psychiatry* 1994;51: 8–19.
Longabaugh R, Woolard RE, Nirenberg TD, et al. Evaluating the effects of a brief motivational intervention for injured drinkers in the emergency department. *J Stud Alcohol* 2001;62:806–816.
Lubelczyk RA, Friedmann PD, Lemon SC, et al. HIV prevention services in correctional drug treatment programs: do they change risk behaviors? *AIDS Educ Prev* 2002;14:117–125.
Macleod J, Oakes R, Copello A, et al. Psychological and social sequelae of cannabis and other illicit drug use by young people: A systematic review of longitudinal, general population studies. *Lancet* 2004;363: 1579–1588.
Madras BK, Compton WM, Avula D, et al. 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;99:280–295.
Mahoney MJ. Self-Change: Strategies for Solving Personal Problems. New York: W. W. Norton; 1979.
Manwell LB, Fleming MF, Johnson K, et al. Tobacco, alcohol, and drug use in a primary care sample: 90-day prevalence and associated factors. *J Addict Dis* 1998;17:67–81.
McCabe SE, Cranford JA, West BT. Trends in prescription drug abuse and dependence, co-occurrence with other substance use disorders, and treatment utilization: Results from two national surveys. *Addict Behav* 2008;33:1297–305.
McCambridge J, Strang J. The efficacy of single-session motivational interviewing in reducing drug consumption and perceptions of drug-related risk and harm among young people: Results from a multi-site cluster randomized trial. *Addiction* 2004;99:39–52.
McCambridge J, Strang J. Deterioration over time in effect of Motivational Interviewing in reducing drug consumption and related risk among young people. *Addiction* 2005;100:470–478.
McKay JR, Lynch KG, Shepard DS, et al. The effectiveness of telephone-based continuing care for alcohol and cocaine dependence: 24-month outcomes. *Arch Gen Psychiatry* 2005;62:199–207.
McLellan AT. Have we evaluated addiction treatment correctly? Implications from a chronic care perspective. *Addiction* 2002;97:249–252.
McLellan AT, Hagan TA, Levine M, et al. Does clinical case management improve outpatient addiction treatment. *Drug Alcohol Depend* 1999;55: 91–103.
McLellan AT, Lewis DC, O'Brien CP, et al. Drug dependence, a chronic medical illness: Implications for treatment, insurance, and outcomes evaluation. *JAMA* 2000;284:1689–1695.
McQueen J, Howe TE, Allan L, et al. Brief interventions for heavy alcohol users admitted to general hospital wards (review). *Cochrane Database Syst Rev* 2009;3:CD005191.
Mertens JR, Lu YW, Parthasarathy S, et al. Medical and psychiatric conditions of alcohol and drug treatment patients in an HMO: Comparison with matched controls. *Arch Intern Med* 2003;163:2511–2517.
Mertens JR, Weisner C, Ray GT, et al. Hazardous drinkers and drug users in

HMO primary care: Prevalence, medical conditions, and costs. *Alcohol Clin Exp Res* 2005;29:989–998.
Miller W. Motivational interviewing with problem drinkers. *Behav Psychother* 1983;1:147–172.
Miller WR, Moyers TB, Arciniega L, et al. Training, supervision and quality monitoring of the COMBINE Study behavioral interventions. *J Stud Alcohol Suppl* 2005;188–195.
Miller WR, Rollnick S. Motivational Interviewing: Preparing People to Change Addictive Behavior. New York: Guilford Press; 1991.
Moyer A, Finney JW, Swearingen CE, et al. Brief interventions for alcohol problems: A meta-analytic review of controlled investigations in treatment-seeking and non-treatment-seeking populations. *Addiction* 2002; 97:279–292.
Newcombe DA, Humeniuk RE, Ali R. Validation of the World Health Organization Alcohol, Smoking and Substance Involvement Screening Test (ASSIST): Report of results from the Australian site. *Drug Alcohol Rev* 2005;24:217–226.
Nilsen P, Baird J, Mello MJ, et al. A systematic review of emergency care brief alcohol interventions for injury patients. *J Subst Abuse Treat* 2008;35:184–201.
Ockene JK, Adams A, Hurley TG, et al. Brief physician- and nurse practitioner-delivered counseling for high-risk drinkers: Does it work? *Arch Intern Med* 1999;159:2198–2205.
Office of National Drug Control Policy. The Economic Costs of Drug Abuse in the United States, 1992–2002. Publication No. 207303. Washington, DC: Executive Office of the President; 2004.
Olfson M, Shea S, Feder A, et al. Prevalence of anxiety, depression, and substance use disorders in an urban general medicine practice. *Arch Fam Med* 2000;9:876–883.
Pasternak AV, Fleming MF. Prevalence of gambling disorders in a primary care setting. *Arch Fam Med* 1999;8:515–520.
Peterson PL, Baer JS, Wells EA, et al. Short-term effects of a brief motivational intervention to reduce alcohol and drug risk among homeless adolescents. *Psychol Addict Behav* 2006;20:254–264.
Pleis JR, Lethbridge-Çejku M. Summary Health Statistics for U.S. Adults: National Health Interview Survey. Vital Health Statistics Series 10, No. 235. Hyattsville, MD: National Center for Health Statistics; 2007. Available at: www.cdc.gov/nchs/data/series/sr_10/sr10_235.pdf. Accessed January 26, 2010.
Pletcher MJ, Kiefe CI, Sidney S, et al. Cocaine and coronary calcification in young adults: the Coronary Artery Risk Development in Young Adults (CARDIA) Study. *Am Heart J* 2005;150:921–926.
Prochaska J, DiClemente C. Toward a comprehensive model of change. In: Miller W, Heather N, eds. Treating Addictive Behaviors: Processes of Change. New York: Plenum Press; 1986.
Raj A, Saitz R, Cheng DM, et al. Associations between alcohol, heroin, and cocaine use and high risk sexual behaviors among detoxification patients. *Am J Drug Alcohol Abuse* 2007;33:169–178.
Regier DA, Farmer ME, Rae DS, et al. Comorbidity of mental disorders with alcohol and other drug abuse. Results from the Epidemiologic Catchment Area (ECA) Study. *JAMA* 1990;264:2511–2518.
Richmond R, Heather N, Wodak A, et al. Controlled evaluation of a general practice-based brief intervention for excessive drinking. *Addiction* 1995;90:119–132.
Rogers CR. The necessary and sufficient conditions of therapeutic personality change. *J Consult Psychol* 1957;21:95–103.
Rollnick S, Heather N, Bell A. Negotiating behavior change in medical settings: The development of brief motivational interviewing. *J Ment Health* 1992;1:25–37.
Rose JS, Herman DS, Hagerty C, et al. Marijuana use among young women in a primary care setting. *J Gen Intern Med* 2007;22:826–829.
Saitz R, Horton NJ, Larson MJ, et al. Primary medical care and reductions in addiction severity: A prospective cohort study. *Addiction* 2005;100: 70–78.
Saitz R, Horton NJ, Sullivan LM, et al. Addressing alcohol problems in primary care: A cluster randomized, controlled trial of a systems intervention. The screening and intervention in primary care (SIP) study. *Ann Intern Med* 2003;138:372–382.
Saitz R, Palfai T, Cheng D, et al. Brief intervention for medical inpatients with unhealthy alcohol use: A randomized, controlled trial. *Ann Intern Med* 2007;146:167–176.
Saitz R, Svikis D, D'Onofrio G, et al. Challenges applying alcohol brief intervention in diverse practice settings: Populations, outcomes, and costs. *Alcohol Clin Exp Res* 2006;30:332–338.
Samet JH, Larson MJ, Horton NJ, et al. Linking alcohol- and drug-dependent adults to primary medical care: A randomized controlled trial of a multi-disciplinary health intervention in a detoxification unit. *Addiction* 2003;98:509–516.
Samet JH, Saitz R, Larson MJ. A case for enhanced linkage of substance abusers to primary medical care. *Subst Abus* 1996;17:181–192.
Skinner HA. The drug abuse screening test. *Addict Behav* 1982;7:363–371.
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: 1155–1160.
Sobell MB, Sobell LC. Problem Drinkers: Guided Self-Change Treatment. New York: Guilford Press; 1993.
Srisurapanont M, Sombatmai S, Boripuntakul T. Brief intervention for students with methamphetamine use disorders: A randomized controlled trial. *Am J Addict* 2007;16:111–116.
Stephens RS, Roffman RA, Curtin L. Comparison of extended versus brief treatments for marijuana use. *J Consult Clin Psychol* 2000;68: 898–908.
Substance Abuse and Mental Health Services Administration. Screening, Brief Intervention, and Referral to Treatment; 2007a. Available at: http://sbirt.samhsa.gov. Accessed January 26, 2010.
Substance Abuse and Mental Health Services Administration. Results from the 2006 National Survey on Drug Use and Health: National Findings. NSDUH Series H-32, DHHS Publication No. SMA 07–4293. Rockville, MD: Office of Applied Studies; 2007b.
Tait RJ, Hulse GK, Robertson SI. Effectiveness of a brief-intervention and continuity of care in enhancing attendance for treatment by adolescent substance users. *Drug Alcohol Depend* 2004;74:289–296.
Tait RJ, Hulse GK, Robertson SI, et al. Emergency department-based intervention with adolescent substance users: 12-month outcomes. *Drug Alcohol Depend* 2005;79:359–363.
Tetrault JM, Crothers K, Moore BA, et al. Effects of marijuana smoking on pulmonary function and respiratory complications: A systematic review. *Arch Intern Med* 2007;167:221–228.
US Preventive Services Task Force. Screening and behavioral counseling interventions in primary care to reduce alcohol misuse: Recommendation statement. *Ann Intern Med* 2004;140:554–556.
US Preventive Services Task Force. Screening for Illicit Drug Use: US Preventive Services Task Force Recommendation Statement. AHRQ Publication No. No. 08–05108-EF-3, January 2008. Agency for Healthcare Research and Quality, Rockville, MD. Available at: http://www.ahrq.gov/clinic/uspstf08/druguse/drugrs.htm. Accessed January 26, 2010.
Voshaar RC, Couvée JE, van Balkom AJ, et al. Strategies for discontinuing long-term benzodiazepine use: Meta-analysis. *Br J Psychiatry* 2006; 189:213–220.
Wenrich MD, Curtis JR, Carline JD, et al. HIV risk screening in the primary care setting. Assessment of physicians skills. *J Gen Intern Med* 1997; 12:107–113.
White HR, Morgan TJ, Pugh LA, et al. Evaluating two brief substance-use interventions for mandated college students. *J Stud Alcohol* 2006;67: 309–317.
Whitlock EP, Polen MR, Green CA, et al. Behavioral counseling interventions in primary care to reduce risky/harmful alcohol use by adults: A summary of the evidence for the U.S. Preventive Services Task Force. *Ann Intern Med* 2004;140:557–568.
Wilk AI, Jensen NM, Havighurst TC. Meta-analysis of randomized control trials addressing brief interventions in heavy alcohol drinkers. *J Gen Intern Med* 1997;12:274–283.
Winters KC, Leitten W. Brief intervention for drug-abusing adolescents in a school setting. *Psychol Addict Behav* 2007;21:249–254.
Woody GE, Gallop R, Luborsky L, et al. HIV risk reduction in the National Institute on Drug Abuse Cocaine Collaborative Treatment Study. *J Acquir Immune Defic Syndr* 2003;33:82–87.
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;104:109–117.
Ziedonis D, Brady K. Dual diagnosis in primary care. Detecting and treating both the addiction and mental illness. *Med Clin North Am* 1997;81: 1017–1036.

Adolescent Alcohol Use and Violence
Are Brief Interventions the Answer?

Richard Saitz, MD, MPH
Timothy S. Naimi, MD, MPH

VIOLENCE AND ALCOHOL USE ARE RISK FACTORS for the 3 leading causes of death among individuals aged 12 to 20 years: unintentional injuries, homicide, and suicide.[1] Thus, the devastating health effects of alcohol and violence on youth lead to an appropriately overwhelming desire to intervene in clinical practice.

Brief counseling could be an answer. In adults, such interventions reduce drinking among those with nondependent unhealthy alcohol use who are identified by screening in primary care settings.[2] These brief counseling interventions are among the most cost-effective but least performed preventive services; are recommended by professional groups, including those that require the highest levels of evidence; and are now reimbursable services.[2,3]

However, the effects of alcohol brief interventions in other settings (eg, emergency departments, hospitals), among adolescents, and on any outcomes besides consumption are less certain.[2,4,5] The US Preventive Services Task Force has found insufficient evidence to recommend for or against screening and counseling to prevent or reduce alcohol misuse by adolescents. Even less evidence is available to support screening and intervention for violence, and results in adults are not promising.[6]

Systematic reviews of studies of alcohol brief interventions for adults in emergency departments find substantial heterogeneity in designs and results, with only about half of studies finding beneficial effects on consumption or consequences.[4,5] Although level I trauma centers are required by accreditation standards to provide alcohol screening and brief intervention, the evidence for efficacy in that setting is limited. One randomized trial had substantial loss to follow-up and a nonsignificant effect on the primary outcome of reinjury. Of 3 subsequent randomized trials, 2 were negative and 1 was positive for reducing the incidence of any arrest for driving under the influence of alcohol in secondary adjusted analyses, but not in primary unadjusted analyses.[4,5,7]

Even less convincing evidence is available for the benefits of brief interventions for alcohol use or violence among youth. For alcohol, 3 randomized trials have tested brief intervention after screening in emergency departments among young people, and results have been inconsistent. One trial found a decrease in drinking but not alcohol consequences, while another found a decrease in consequences but not in drinking.[8,9] A third trial (http://www.clinicaltrials.gov; identifier: NCT00183157) has recently been completed, but the results have not yet been published. The prevention of youth violence has not been sufficiently studied in emergency settings,[10] and the results of the few available studies have been mixed. One trial reported a reduction in self-reported (but not trauma registry–documented) reinjury,[11] suggesting that self-reported outcomes for violence may be biased in this population.

Despite the lack of evidence for the effectiveness of interventions to reduce alcohol use and violence separately, it does seem logical to address these risky behaviors together, because they so often co-occur. This was the approach taken by Walton et al in this issue of *JAMA*.[12] Adolescents receiving emergency services for illness or injury who reported both alcohol use and physical aggression were randomized to receive a brochure, a 35-minute motivational intervention delivered by a therapist, or a self-administered animated computer intervention.

Overall, positive results were noted for few outcomes, effect sizes were small, and none of the observed 3-month benefits were sustained at 6 months. The 24

See also p 527.

Author Affiliations: Clinical Addiction Research and Education Unit, Section of General Internal Medicine, Boston Medical Center, and Boston University School of Medicine (Drs Saitz and Naimi), and Department of Epidemiology, Boston University School of Public Health (Dr Saitz), Boston, Massachusetts.
Corresponding Author: Richard Saitz, MD, MPH, Section of General Internal Medicine, Boston Medical Center, 801 Massachusetts Ave, Second Floor, Boston, MA 02118 (rsaitz@bu.edu).

comparisons of binary outcomes (2 intervention groups compared with controls, 2 outcome time points, 3 alcohol and 3 violence outcomes) did not show statistically significant group differences in "alcohol misuse" (ie, unhealthy alcohol use) or "binge drinking" (ie, ≥5 drinks) at 3 or 6 months; however, both intervention groups had significantly larger (by 6%-8%) decreases in alcohol-related consequences vs the control group at 6 (but not 3) months. Computer-delivered interventions did not affect any of the violence-related outcomes except being the recipient of peer violence at 6 months. Therapist counseling had a modest effect (eg, 13% absolute reduction in peer aggression compared with controls) for all 3 violence-related outcomes at 3 months but not at 6 months. Of the additional 24 comparisons of frequency-based outcomes, only 1 was statistically significant: a lower number of violence consequences at 3 months in the therapist group.

In addition to the mixed results and modest effects, the study's trial registration suggests that additional primary outcomes were measured, such as drug use, injury, delinquency, and weapon carrying,[13] raising the concerns that type I error in the setting of multiple comparisons (7 of 48 reported comparisons were statistically significant at a 1-tailed $P < .05$) or a focus on more positive outcome domains might explain the results. If this study had measured more objective outcomes such as physician-documented injury events or school-based reports of violent incidents, rather than self-reported risk behaviors, the findings might have been more convincing. Although the authors used validated survey instruments, social desirability bias is of concern with self-reported outcomes, because study participants in the counseling groups might be less likely to report behaviors in follow-up that they were told were unsafe or undesirable. This might be a greater concern for study participants who received the therapist intervention than those who received the computer intervention or the brochure. Violence and alcohol consequences might be seen by adolescents as most embarrassing, and these accounted for the observed differences, whereas behaviors that were possibly perceived as less embarrassing, such as alcohol use, were not less reported in the intervention groups. Taken together, these concerns support the authors' recommendation that their study should be replicated in other settings.

Even a modest benefit of a relatively inexpensive intervention for common health problems might be worth pursuing on a large scale. However, there are barriers to the widespread implementation of screening and brief intervention programs in clinical settings. Although brief intervention for alcohol has been recommended for decades in primary care settings, levels of implementation have been dismal for a variety of reasons, such as inadequate training, lack of clinician time, and inadequate reimbursement.[3] Although computerized intervention seems more likely to be disseminated successfully than a 35-minute expert therapist–delivered session, the computer intervention used in the study by Walton et al[12] affected only 1 secondary violence outcome (a decrease in being a recipient of peer violence) at 1 time point. However, the computer intervention was associated with a decrease in alcohol consequences, which is a potentially promising use for this intervention.

The most proven and effective method to reduce youth drinking, and likely alcohol-related violence, is to implement population-based strategies such as raising alcohol excise taxes and enforcing minimum legal drinking age laws.[14] These strategies have been neglected. For example, the federal beer tax, which is based on a fixed amount per volume, has eroded by almost 40% in real terms since it was last adjusted in 1991. Until the findings of Walton et al[12] can be replicated (and hopefully improved), brief interventions for violence prevention in emergency departments do not seem promising. However, existing evidence supports the implementation of screening and brief intervention for unhealthy alcohol use in adult primary care settings, especially for young adults in whom the prevalence of this risky behavior is greatest.

For adolescents, alcohol screening is currently recommended by the American Academy of Pediatrics.[15] While specific screening strategies for underage youth have not been fully elucidated, the focus of screening should be on any use of alcohol (with appropriate assessment of those screening positive), irrespective of their experience with violence, because any use of alcohol by adolescents, not just binge drinking, is associated with a variety of adverse outcomes.[16] However, the ultimate benefit of such screening remains dependent on the development, testing, and implementation of effective clinical strategies to reduce youth alcohol consumption and violence and their adverse consequences.

Financial Disclosures: Dr Saitz reports having been a consultant for online alcohol-related screening and brief intervention education projects supported by National Institutes of Health (NIH) grants to Medical Directions and Inflexxion, and for NIH grants to Kaiser Permanente, the University of Massachusetts, and Brandeis University. He also has been compensated by Beth Israel Deaconess Hospital and the National Institute on Alcohol Abuse and Alcoholism for serving on data and safety monitoring boards. He is compensated for editorial work by the Massachusetts Medical Society, the British Medical Journal Group, and the American Society of Addiction Medicine. He has developed educational materials for Fusion medical education and consulted for Saatchi and Saatchi Healthcare on alcohol dependence treatment. He has or anticipates being compensated as a speaker on alcohol and drug topics by multiple government agencies, academic institutions, and professional societies. He has also provided expert opinion on legal cases involving identification and management of alcohol and drug-related problems. Dr Naimi reports currently receiving NIH grant support and was supported through his federal salary while employed by the US Public Health Service.

REFERENCES

1. Centers for Disease Control and Prevention National Center for Injury Prevention and Control. WISQARS leading causes of death reports, 1999-2007. http://webappa.cdc.gov/sasweb/ncipc/leadcaus10.html. Accessed July 1, 2010.

2. US Preventive Services Task Force. Screening and behavioral counseling interventions in primary care to reduce alcohol misuse. April 2004. http://www.ahrq.gov/clinic/uspstf/uspsdrin.htm. Accessed July 12, 2010.
3. 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):143-152.
4. Saitz R. Candidate performance measures for screening for, assessing, and treating unhealthy substance use in hospitals: advocacy or evidence-based practice? *Ann Intern Med*. 2010;153(1):40-43.
5. Bernstein E, Bernstein JA, Stein JB, Saitz R. SBIRT in emergency care settings: are we ready to take it to scale? *Acad Emerg Med*. 2009;16(11):1072-1077.
6. MacMillan HL, Wathen CN, Jamieson E, et al; McMaster Violence Against Women Research Group. Screening for intimate partner violence in health care settings: a randomized trial. *JAMA*. 2009;302(5):493-501.
7. Gentilello LM, Rivara FP, Donovan DM, et al. Alcohol interventions in a trauma center as a means of reducing the risk of injury recurrence. *Ann Surg*. 1999;230(4):473-480.
8. Monti PM, Colby SM, Barnett NP, et al. Brief intervention for harm reduction with alcohol-positive older adolescents in a hospital emergency department. *J Consult Clin Psychol*. 1999;67(6):989-994.
9. Spirito A, Monti PM, Barnett NP, et al. A randomized clinical trial of a brief motivational intervention for alcohol-positive adolescents treated in an emergency department. *J Pediatr*. 2004;145(3):396-402.
10. Snider C, Lee J. Youth violence secondary prevention initiatives in emergency departments: a systematic review. *CJEM*. 2009;11(2):161-168.
11. Zun LS, Downey L, Rosen J. The effectiveness of an ED-based violence prevention program. *Am J Emerg Med*. 2006;24(1):8-13.
12. Walton MA, Chermack ST, Shope JT, et al. Effects of a brief intervention for reducing violence and alcohol misuse among adolescents: a randomized controlled trial. *JAMA*. 2010;304(5):527-535.
13. Tailored teen alcohol and violence prevention in the emergency room (ER). http://clinicaltrials.gov/ct2/show/NCT00251212. Accessed July 2, 1010.
14. Bonnie RJ, O'Connor ME, eds; Committee on Developing a Strategy to Reduce and Prevent Underage Drinking; National Research Council; Institute of Medicine. *Reducing Underage Drinking: A Collective Responsibility*. Washington, DC: National Academies Press; 2004.
15. Kokotailo PK; Committee on Substance Abuse. Alcohol use by youth and adolescents: a pediatric concern. *Pediatrics*. 2010;125(5):1078-1087.
16. 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.

Intimate Partner Violence Against Women
What Outcomes Are Meaningful?

Angela J. Taft, PhD
Kelsey L. Hegarty, MBBS, PhD

Intimate partner violence (IPV) is estimated to be the leading contributor to the global burden of mental health problems among women of reproductive age.[1] There is an increasing urgency for rigorous, good-quality evidence about what is effective in preventing or ameliorating such harm in community and health care settings.[2] A recent Cochrane review of partner violence advocacy trials found only 2 trials conducted in community settings and overall concluded that evidence of health benefit is scarce in any setting.[3]

In this issue of *JAMA*, Tiwari and colleagues[4] report a study in which Chinese women survivors of IPV attending a Hong Kong multipurpose community center were randomly assigned to receive a 12-week advocacy intervention to reduce depression and IPV compared with usual community services. The intervention consisted of a 30-minute empowerment session delivered by a registered social worker (trained to implement a social support and empowerment protocol) followed by weekly telephone social support from the same social worker. The empowerment session involved a focus on enhanced choice making and problem solving, combined with safety assessment and planning. Staff providing telephone support reminded women that help was available and aimed to respond flexibly to the expressed stressors and needs in women's lives at the time.

See also p 536.

The 200 women participating in this study were married, in their late 30s, and living in secure housing in a stable community in which they remained for the duration of the study. Their abuse status was assessed with the Chinese version of the Revised Conflict Tactics Scale (C-CTS2) and their depressive symptoms with the Chinese version of the Beck Depression Inventory II (C-BDI-II), with measurements obtained at baseline, at 3 months (when the intervention ended), and at 9 months.[4] Although mean scores for partner psychological aggression as measured by the C-CTS2 (eg, shouted at, stomped out of room) were elevated at baseline, mean scores for physical assault and sexual coercion were low and remained low throughout the study. With the exception of 2 women, none had disclosed any abuse previously or sought help from health or social services. Women in both study groups reported mean rates of severe depression at baseline, moderate depression at 3 months, and mild depression at 9 months. Women in both groups reported an initial increase in mean rates of partner psychological aggression and then low rates at study completion. The reported mean differences between the study groups on both measures were modest but did reach statistical significance. However, the difference of –2.7 units in depression scores did not reach the clinically meaningful level of 5 units that is recommended for the C-BDI-II.[5] In the follow-up telephone support calls, women's expressed needs focused on

Author Affiliations: Mother and Child Health Research, La Trobe University (Dr Taft); and University of Melbourne (Dr Hegarty), Melbourne, Victoria, Australia.
Corresponding Author: Angela J. Taft, PhD, Mother and Child Health Reasearch, La Trobe University, 324-328 Little Lonsdale St, Melbourne, VIC, Australia 3000 (a.taft@latrobe.edu.au).

Interventions Targeting HIV-Infected Risky Drinkers

Drops in the Bottle

Jeffrey H. Samet, M.D., M.A., M.P.H., and Alexander Y. Walley, M.D., M.Sc.

Alcohol use is common among people infected with HIV and may contribute to adverse consequences such as reduced adherence to treatment regimens and increased likelihood of risky sexual behaviors. Therefore, researchers and clinicians are looking for treatment approaches to reduce harmful alcohol consumption in this population. However, clinical trials of existing treatment models are scarce. A literature review identified only 11 studies that included HIV-infected patients with past or current risky alcohol use and which targeted alcohol use and other health behaviors. Four studies focusing on HIV-infected participants with alcohol problems found mixed effects on adherence and on alcohol use. Five clinical trials included at least 10 percent of HIV-infected subjects who use alcohol; of these, only one reported significant evidence of a favorable impact on alcohol consumption. Finally, two trials targeting alcohol users at high risk for HIV infection identified treatment effects that were not sustained. Taken together, these findings provide limited evidence of the benefit of behavioral interventions in this population. Nevertheless, these studies give some guidance for future interventions in HIV-infected patients with alcohol problems. *Key Words: Alcohol and other drug use; alcohol consumption; alcohol use disorder; human immunodeficiency virus; HIV-infected patients; sexually transmitted disease; unsafe sex; treatment method; treatment outcome; intervention; clinical trial; literature review*

In the United States, people infected with the human immunodeficiency virus (HIV) drink more alcohol than people in the general population. Specifically, a higher proportion drink risky amounts[1] or have an alcohol use disorder (i.e., abuse or dependence) (Conigliaro et al. 2003; Galvan et al. 2002; Lefevre et al. 1995; Samet et al. 2003*a,b*, 2004). Risky alcohol use in HIV-infected people has been associated with the following range of adverse effects:

- Reduced adherence to medication regimens for treatment of HIV infection (Chander et al. 2006; Conen et al. 2009; Cook et al. 2001; Golin et al. 2002; Halkitis et al. 2003; Samet et al. 2004);
- Lack of a health care provider for the HIV infection (Metsch et al. 2009);
- Delayed linkage to HIV medical care (Samet et al. 1998);
- Increase in risky sexual behaviors (Kalichman et al. 2002; Metsch et al. 2009);
- Increased transmission of sexually transmitted infections (Kalichman et al. 2000); and
- Progression of HIV disease (Conigliaro et al. 2003; Miguez et al. 2003; Samet et al. 2007).

Given the spectrum of problems associated with such alcohol use among HIV-infected patients, one important avenue to improving the health of this population is to develop interventions

[1] According to the National Institute on Alcohol Abuse and Alcoholism (2007), women who drink more than 3 drinks per day or more than 7 drinks per week and men who drink more than 4 drinks per day or more than 14 drinks per week are at increased risk for alcohol-related problems. Alcohol consumption levels above these limits are considered risky drinking.

Jeffrey H. Samet, M.D., M.A., M.P.H., *is a professor in the Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, and in the Department of Social and Behavioral Sciences, Boston University School of Public Health, both in Boston, Massachusetts.*

Alexander Y. Walley, M.D., M.Sc., *is an assistant professor in the CARE Unit, Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, Boston, Massachusetts.*

that target alcohol use and its associated consequences. Accordingly, interventions have been designed to both decrease alcohol consumption and address the specific adverse health consequences.

The concept that negative consequences of alcohol use can be reduced in patients with HIV infection is based on research demonstrating the impact of clinical interventions on alcohol consumption and associated negative consequences in patients without HIV infection (Institute of Medicine 1990; Kristenson et al. 1983). Alcohol research over the past three decades has demonstrated that behavioral interventions can be effective, with benefits varying based on setting, severity of alcohol problems, and patient characteristics. For example, meta-analyses of randomized controlled trials (RCTs)[2] of interventions to reduce risky alcohol use demonstrated decreased drinking for patients in primary care settings (Beich et al. 2003; Kaner et al. 2007). However, no such effects were found in meta-analyses of interventions delivered in hospital settings (Emmen et al. 2004), possibly because inpatients typically have greater severity of alcohol problems (i.e., most are alcohol dependent) (Saitz et al. 2007, 2008). Several high-quality RCTs of brief interventions delivered in emergency departments also detected no or limited benefit (D'Onofrio and Degutis 2002; Daeppen et al. 2007; Longabaugh et al. 2001; Monti et al. 1999). The influence of the patient's consumption levels also was demonstrated in several studies. For example, in two separate RCTs in the primary-care setting (Fleming et al. 1997; Ockene et al. 1999), where patients were seeking medical care but not necessarily for an alcohol problem, implementation of a 5- to 15-minute discussion reduced alcohol consumption in patients who met the criteria for risky drinking. Studies of such brief interventions among patients who met the criteria for alcohol dependence, however, have shown no benefit (Kaner et al. 2007; Whitlock et al. 2004; Wutzke et al. 2002).

For alcohol-dependent patients, more extensive behavioral interventions (e.g., cognitive–behavioral coping skills, motivational enhancement, 12-step facilitation) can be effective (*Project MATCH* Research Group 1997). In addition, several medications (i.e., disulfiram, naltrexone, and acamprosate) are approved for the treatment of alcohol dependence, and other medications (e.g., topiramate) are being further evaluated (Anton et al. 2006; Garbutt et al. 2005; Kranzler and Van Kirk 2001; Olmsted and Kockler 2008; Rubio et al. 2001).

Given the strong evidence that alcohol consumption is an important health issue for many people with HIV infection, efforts to potentially ameliorate these problems by addressing alcohol use are of great interest. The studies in non–HIV-infected people reviewed above suggest that interventions among HIV-infected people with alcohol problems could be beneficial. However, the wide range of results in these intervention studies based on setting and disease severity argues for the need to carefully assess efforts to mitigate alcohol's deleterious impact on health in HIV-infected patients. As an important step in this direction, this article summarizes the findings of a review of the clinical trial literature on interventions addressing alcohol consumption and its consequences among HIV-infected patients. After describing the design of the literature search and evaluation, the article reviews the findings of the studies identified and discusses the implications of those findings.

Design of the Literature Review

The literature review sought to identify clinical trials of interventions among HIV-infected people with past or current unhealthy alcohol use (i.e., the spectrum from risky drinking to alcohol dependence [Saitz 2005]) that reported effects on any of the following outcomes:

- HIV disease progression;
- Receipt of HIV treatment;
- HIV medication adherence;
- HIV risk behaviors;
- Acquisition of sexually transmitted infections; and
- Alcohol use.

To be included in the review, the studies had to report alcohol-specific outcomes. Beyond that, the studies were classified into three categories of specificity. The most specific category comprised clinical trials that included only HIV-infected people with past or current unhealthy alcohol use. The second category comprised clinical trials that included only HIV-infected people but in which not all of the participants exhibited unhealthy alcohol use. For a study to be included in this category, at least 10 percent of participants had to report current alcohol use. The third category of studies comprised trials that were aimed at preventing alcohol use and sexual behaviors that put people at risk of HIV infection among alcohol-using people. Although these studies did not include HIV-infected participants or did not report the HIV status of the participants, they were reviewed because they may inform future research on people at risk of HIV transmission in the setting of alcohol use.

Initially, the review intended to include only RCTs. However, very few studies were identified that met this criterion in the first two categories. Therefore, the search was expanded to include nonrandomized and non-controlled clinical intervention trials in categories 1 and 2.

To identify relevant studies, the literature database MEDLINE was searched through September 30, 2009, using the search terms "HIV, alcohol, hazardous drinking, risky drinking, problem drinking, counseling, brief intervention, 12 step, pharmacotherapy, naltrexone, acamprosate, disulfiram, topiramate, and clinical trial." For all articles identified using this approach, the reference lists also

[2] RCTs are clinical studies in which patients randomly are assigned to either one or more groups receiving the treatment under investigation or to a control group receiving no treatment or a treatment of known efficacy.

were scanned, as were related articles identified by the search engine for the MEDLINE data base to look for additional studies. Reference lists for articles that were closely related, but did not meet the criteria, also were reviewed. Finally, articles referenced in relevant review articles were examined. Titles of all articles were reviewed to determine if the articles met the selection criteria. If the nature of the study could not be discerned through the title, the abstract and/or full text of the article was retrieved and reviewed.

For all studies that met the criteria for one of the three categories, information on the setting, study design, methodological quality, type of intervention, outcomes reported, period of follow-up, and results was extracted. The following sections summarize the findings of these analyses. They are presented as a descriptive narrative synthesis because studies were too few and heterogeneous to perform a standard meta-analysis.

Results of the Literature Review

The search strategy described above identified 241 potentially relevant studies that were evaluated further. Of these, four studies including a total of 578 patients (Aharonovich et al. 2006; Parsons et al. 2007; Samet et al. 2005; Velasquez et al. 2009) met the selection criteria for the first category (see table 1). Another five clinical trials that included 1,311 patients (Gilbert et al. 2008; Naar-King et al. 2006, 2008; Rotheram-Borus et al. 2001, 2009; Sorensen et al. 2003) fell into the second category. In addition, two informative studies of interventions among people at high-risk for HIV reported outcomes specific to alcohol use (Kalichman et al. 2008; Morgenstern et al. 2007). All of these studies are reviewed below. Some other studies that involved alcohol-using, HIV-infected patients, but were excluded from this discussion because of serious design or methodological limitations, are listed in Table 2 because they may inform additional research. Interestingly, no controlled trials of the four medications recommended by NIAAA (2007) for the treatment of alcohol dependence (i.e., disulfiram, naltrexone, acamprosate, and topiramate) have been conducted in HIV-infected patients.

Clinical Trials Among HIV-Infected People With Past or Current Unhealthy Alcohol Use

Velasquez and Colleagues (2009) Study. These investigators conducted an RCT among 253 HIV-infected men who had had sex with men in the previous 3 months and who scored more than eight points on the AUDIT questionnaire (Babor et al. 2001). The intervention group received four manual-guided individual sessions and four manual-guided peer education and support group sessions that utilized motivational interviewing (MI) counseling strategies (Miller and Rollnick 2002) to guide participants through the stages of change of Prochaska and DiClemente's Trans-Theoretical Model[3] (Prochaska and DiClemente 1982). In contrast, the control group received educational materials on HIV and alcohol, referral information, and advice to stop or cut back on their alcohol use. At the 12-month follow-up, the investigators determined some benefits of the intervention on some of the measures evaluated. For example, the control group had 1.4 times the number of drinks per 30 days and 1.5 times the number of heavy-drinking days per 30 days compared with the intervention group. For other measures (e.g., having anal sex without a condom, number of drinking days, or number of days on which both drinking and sex occurred), however, no significant difference existed between the two groups. Only when the analysis of same-day drinking and sex was restricted to participants who had shown this behavior at baseline, did those in the control group have significantly (i.e., 2.19 times) more days on which drinking and sex occurred than the intervention group. The interpretation of these findings is limited by the fact that there was differential loss to follow-up—that is, the analyses included only 81 percent of participants randomized to the intervention group and 90 percent of subjects randomized to the control group. Thus, one cannot exclude the possibility that particularly in the intervention group, participants with worse outcomes were not included in the analysis.

Aharonovich and Colleagues (2006) Study. In this pilot study, 31 HIV-infected primary-care patients with heavy alcohol use received one session of MI from a trained counselor, followed by daily telephone-based interactive voice response (IVR) assessments of drinking amounts and graphic feedback of changes in drinking at 30 and 60 days. This intervention resulted in a decrease in the number of drinks per day at 30 and 60 days (from 3.2 drinks per day at baseline to 1.7 drinks at 30 days and 1.2 drinks at 60 days). The IVR system was utilized; 77 percent of all possible daily calls were completed at 30 days. However, these improvements can not be attributed to the intervention with confidence because there was no control group.

Parsons and Colleagues (2007) Study. These investigators conducted an RCT among 143 HIV-infected people with "hazardous drinking" (defined as more than 16 standard drinks per week for men or more than 12 standard drinks per week for women), assessing treatment effects on HIV medication adherence and alcohol outcomes. The intervention involved eight 1-hour individual sessions of MI and cognitive behavioral skills training over 3 months and was compared with a time- and content-equivalent control.[4] Over the

[3] The transtheoretical model (TTM) is a health behavior theory that assesses the individual's readiness to change a particular behavior in order to facilitate the desired behavior change. The stages of change are: precontemplation, contemplation, preparation, action, and maintenance.

[4] With a time- and content-equivalent control group, participants in that group spend the same amount of time with a health care provider/therapist as the intervention group, and they receive the same type of information. The only difference between the intervention and control groups is the method used to deliver the information, allowing researchers to determine whether one approach is more effective than the other.

follow-up period (3 and 6 months), both groups exhibited substantial improvement for both total alcohol drinks over 14 days or drinks per drinking day, although no significant differences existed between the intervention and the control group. However, compared with the control group, the intervention did improve medication adherence, number of virus particles detectable in the blood (i.e., viral load), and CD4 cell[5] counts at 3 months. These statistically significant improvements were not sustained at 6 months.

[5] CD4 cells are a type of white blood cell that is the main target of the HIV virus; accordingly, levels of these cells in the blood decline with progressing HIV infection and are a marker for disease progression.

Samet and Colleagues (2005) Study. This RCT included 151 HIV-infected patients on antiretroviral therapy (ART) who had a history of alcohol problems. The participants received either four nurse-delivered, 30- to 60-minute sessions focusing on HIV medication adherence and alcohol counseling, both in a clinic

Table 1 Studies Identified During a Literature Search on Interventions to Decrease Alcohol Use and Related Behaviors among HIV-Infected People and Alcohol Users at High Risk for Infection

Study	Population/Setting	Design	Outcomes/Results	Comments
Category 1: Clinical trials among HIV-infected people with past or current unhealthy alcohol use				
Velasquez et al. 2009	**Population**: 253 HIV-infected men who had sex with men (MSM) in the previous 3 months and an AUDIT score of more than 8. **Setting**: Recruited from HIV organizations, advertising, and social venues between 1999 and 2003.	**Intervention:** Randomized Controlled Trial (RCT) of four sessions of motivational interviewing (MI)-based individual counseling and four sessions of transtheoretical model–based peer-group education/support. **Control:** HIV and alcohol educational materials, resource referrals, and advice to stop or reduce drinking. **Assessment:** Baseline, 3, 6, 9, and 12 months.	**Alcohol use:** Control group had 1.38 times the number of drinks per 30 days and 1.50 times the number of heavy drinking days per 30 days compared with the intervention group. **Sex risk:** No significant effect was demonstrated for anal sex without a condom or number of days on which drinking and sex occurred.	**Alcohol measures**: AUDIT, 90-day timeline follow-back (TLFB) at follow-up assessments. Differential loss to follow-up at 12 months (34% in intervention group and 26% in control group). Only 95 of 118 (81%) of the intervention group and 121 of 135 (90%) of the control group were included in the analyses.
Aharonovich et al. 2006	**Population**: 31 HIV-infected men and women engaged in HIV primary care. **Alcohol use**: All had four or more drinks at least once in the past 30 days, 55% had five or more drinks in the last week. **Setting**: HIV primary care clinic.	**Intervention**: 30-minute MI session on reducing alcohol use by counselor trained in MI plus an automated daily telephone self-monitoring interactive voice response (IVR) system with graphical feedback at 30-day follow-up meetings. **Control**: No control group. **Assessment:** Baseline, 30, 60, and 90 days.	**Drinks per day**: Using 7-day recall, mean drinks per day was 3.2 at baseline, 1.7 at 30 days, and 1.2 at 60 days. Mean highest drinks per day was 8.4, 4.1, and 3.8, respectively. **Cocaine use:** Decreased significantly at 60 days.	**Alcohol measures**: Quantity and frequency in past week and past month. Qualitative assessment of the program demonstrated satisfaction with daily calling and the feedback graph. Not a randomized controlled trial.
Parsons et al. 2007	**Population**: 143 HIV-infected subjects on antiretroviral therapy (ART) with hazardous drinking (more than 16 drinks per week for men, more than 12 drinks per week for women) recruited through HIV clinics and advertising from 2002 to 2005. **Setting**: Behavioral research center.	**Intervention**: RCT of eight 60-minute MI plus cognitive behavioral skills training (CBST) session by Masters-level counselors. **Control**: Eight 60-minute time and content-equivalent education sessions by health educators. All sessions delivered individually in private office over 12 weeks. **Assessment**: Baseline, 3 and 6 months.	**Alcohol use**: No significant effects on total drinks over 14 days or drinks per drinking day. Decreases in both groups from baseline to 3 and 6 months for these two drinking outcomes. **Medication adherence**: Significant improvement in dose and day adherence at 3 months, but difference not retained at 6 months. **HIV viral load/CD4 cell count**: Significant improvement at 3 months but not at 6 months.	**Alcohol measure**: Self-report 14-day TLFB to calculate total drinks and drinks per drinking day. **Adherence measures**: Self-report dose adherence = number of doses taken/number of doses scheduled over 14 days. Self-report day adherence = number of days with perfect adherence/14 days.

and at home, or no intervention. The study found no significant differences between groups upon examination of the following outcomes: 3-day medication adherence, 30-day adherence, CD4 cell count, viral load, drinks per day, percent reporting drinking, or percent reporting hazardous drinking. Study limitations were that not all participants were non-adherent to their HIV medication at baseline and a substantial percentage were not in the risky-drinking range of unhealthy alcohol use, the group most amenable to brief interventions.

Clinical Trials Among HIV-Infected People of Whom at Least 10 Percent Currently Use Alcohol

Five studies identified in the literature review fell into this category, and only one of these (Rotheram-Borus et al. 2009) demonstrated significant treatment effects on alcohol use (see table 1). This study was a subanalysis of a parent RCT among 936 HIV-infected people who were sexually active without a condom with at least one HIV-negative partner or two HIV-infected partners (Wong et al. 2008). The participants received either 15 90-minute individual sessions of cognitive–behavioral therapy (CBT) delivered over 15 months or usual care. The subanalysis by Rotheram-Borus and colleagues (2009) was limited to 270 HIV-infected participants who were homeless or without stable housing. In this group, the intervention was found to reduce alcohol or marijuana use from 36 to 28 days in the prior 90 days, whereas in the control group the frequency of alcohol or marijuana use was unchanged at 35 of the last 90 days. However, this

Table 1 con't

Study	Population/Setting	Design	Outcomes/Results	Comments
Samet et al. 2005	**Population**: 151 HIV-infected patients on ART, with current or lifetime alcohol problems, determined by two or more positive responses on CAGE questionnaire or clinical diagnosis of alcohol disorder recruited from 1997 to 2000. **Setting**: Hospital (patients receiving HIV medical care).	**Intervention**: RCT of four 15- to 60-minute sessions over 3 months with MI-trained nurse who (1) addressed alcohol problems, (2) educated about ART efficacy, and (3) delivered tailored adherence advice including a reminder watch and a home visit. **Control:** Standard care **Assessment:** Baseline, 6, and 12 months.	**Alcohol use**: No significant effects on drinks per day, percent reporting any drinking, percent reporting hazardous drinking. **Medication adherence**: No significant effects on 3-day or 30-day adherence. **HIV viral load/ CD4 cell count**: No significant effects on mean CD4 cell count or mean log HIV RNA.	**Alcohol measures**: Self-report 30-day alcohol use from the Addiction Severity Index. **Adherence measures**: Self-reported AIDS Clinical Trial Group scale with 100% and 95% or more thresholds at 3-day and 30-day adherence, respectively.
Category 2: Clinical trials among HIV-infected people of whom at least 10% have current alcohol use				
Rotheram-Borus et al. 2009	**Population**: 270 HIV-infected people sexually active without a condom with at least one HIV-negative partner or two HIV-infected partners who were marginally housed and had four or more assessments; recruited from 2000 to 2002. **Alcohol use**: Mean number of days using alcohol or marijuana in the last 90 was 37. **Setting**: Recruited from community agencies, medical clinics, and advertisements.	**Intervention**: RCT of 15 90-minute individual counseling sessions, organized in three modules ("Coping" at 0–5 months, "Act Safe" at 5–10 months, and "Stay Healthy" at 10–15 months). **Control**: No intervention, only assessments **Assessment**: Baseline, 15, 20, and 25 months.	**Alcohol or marijuana use in the last 3 months**: At 25 months, the intervention group reduced its use from 36 to 28 days in the prior 90 days, whereas the control group was unchanged at 35 days of the last 90. Number of HIV negative partners and risky sexual acts also was reduced.	**Subanalysis of a clinical trial (Wong et al. 2008)**: 5% used alcohol/ marijuana in the parent study. Proportion of alcohol users at baseline not presented in this study. Parent study reported only transmission act outcomes and demonstrated an effect that was not maintained at 25 months. Imbalance in transmission risk acts at baseline resulted in ineffective randomization, thus propensity scores were used to adjust for imbalances.

Table 1 con't

Study	Population/Setting	Design	Outcomes/Results	Comments
Naar-King et al. 2006, 2008	**Population**: 65 HIV-infected patients, aged 16–25 regardless of alcohol use or risk behaviors. **Alcohol use**: 77% lifetime, 39% had used alcohol in last 30 days at study entry. **Setting**: Adolescent HIV care clinic within a tertiary care children's hospital.	**Intervention**: RCT of four 60-minute sessions of motivational enhancement. Therapy focused on two of three areas: substance use, sexual risk, or medication adherence over 10 weeks. **Control**: Wait list and standard care. **Assessment:** Baseline, 3, 6, and 9 months.	No significant effects at 9-month follow-up. **Alcohol use**: Borderline significant reduction in number of drinks in the week containing the maximum number of drinks (–9.65 vs. –1.3) at 3 months (n = 51). **Marijuana use**: Borderline significant reduction in number of times marijuana was used (n = 65). **Sexual risk**: Borderline significant reduction in total number of intercourse acts without a condom at 6 months (n = 65). **HIV viral load**: Significant reduction in log viral load at 6 months (n = 65).	**Alcohol and drug measures**: Timeline follow-back, though time window is not stated. **Sex risk measure**: Total number of unprotected intercourse acts without a condom. Note: 3-month outcomes on 51 subjects were published in 2006 and 6- and 9-month outcomes on 65 subjects published in 2008.
Gilbert et al. 2008	**Population**: 476 patients with alcohol risk (38%), defined as exceeding NIAAA safe drinking limits or drug risk (42%), or sex risk (60%), were recruited between 2003 and 2006. **Setting**: Outpatient HIV clinics.	**Intervention**: RCT of two sessions of tailored risk-reduction counseling at study entry and 3 months using a MI "Video Doctor" via laptop computer, printed educational worksheet, and delivery of a cueing sheet on reported risks to clinic care providers. **Control**: Standard care. **Assessment**: Baseline, 3, and 6 months.	**Alcohol use**: No significant effects on any risky drinking or number of drinks per week. **Drug use**: Significantly decreased 30-day illicit drug use at 3 and 6 months and fewer days of illicit drug use at 6 months. **Sex risk**: Significantly decreased 3-month unprotected sex at 3 and 6 months and fewer casual sex partners at 6 months. No effects on condom use.	**Alcohol measures**: Self-reported NIAAA risky drinking over 3 months. **Drug use measures**: Self-reported drug use over 30 days included any cocaine, methamphetamine, or heroin or 3 or more days of barbiturates, prescription opiates, hallucinogens, inhalants, or methylene-dioxymethamphetamine (MDMA).
Sorensen et al. 2003	**Population**: 190 HIV-infected patients with substance dependence; recruited from inpatient medical wards, detoxification clinic, and the emergency department from 1994 to 1996. **Alcohol use**: 61% in the last 30 days. **Setting**: Public general hospital.	**Intervention**: RCT of 12 months of case management by certified substance counselors in the community with caseload of 1:20 **Control**: Single brief contact with education about reducing HIV risk, information on HIV services, referrals to addiction treatment, social services. **Assessment**: Baseline, 6, 12, and 18 months.	No outcomes showed significant change between study groups at any time points, except decreased sex risk index. **Outcomes measured**: Addiction severity index composite scores, AIDS risk assessment scores, Beck depression inventory, health status questionnaire, and support evaluation list.	Summary/index score is shown without explanation of the raw measure.

Table 1 con't

Study	Population/Setting	Design	Outcomes/Results	Comments
Rotheram-Borus et al. 2001	**Population**: 310 HIV-infected patients (age 13–24) from nine adolescent clinics recruited from 1994 to 1996. **Alcohol use**: 67% nonabstinent at baseline. **Setting**: Adolescent clinics.	**Intervention**: 23 group sessions of two modules ("Stay Healthy" and "Act Safe"). **Control**: Standard care. Eligible for receiving the intervention at the conclusion of the study. **Assessment**: Baseline, 9, and 15 months.	**Alcohol/marijuana use:** 63% for attendees vs. 67% for control vs. 84% for nonattendees at 15 months.	Sequential assignment of 15 youths to intervention versus control groups (not randomized). The reported comparisons were attendees versus non-attendees versus control subjects. No intention-to-treat analysis was reported. Differential loss to follow-up. No alcohol-specific outcome was reported.
Category 3: Randomized controlled trials among alcohol users at high risk for HIV infection				
Morgenstern et al. 2007	**Population**: 198 MSM with current alcohol user disorder. **Alcohol use**: 88% with alcohol dependence. Mean drinks per drinking day was 10.4. **Setting**: Subjects recruited through advertisements in gay media, internet chat rooms, outreach to gay bars and clubs.	**Intervention**: 12 sessions of combined MI and coping skills training (MI+CBT) over 12 weeks (n = 47). **Control**: Four sessions of MI over 12 weeks (n = 42). Non–help-seeking (NHS) control group (n = 109). **Assessment**: Baseline, 12 weeks, and 12 months.	**Drinks per day**: At 12 weeks, the MI group had greater decreases in drinks per day than the MI+CBT group. This difference was not sustained at 12 months. Both intervention groups had greater decreases then the NHS group, but the NHS group also had substantial decreases in drinking.	**Alcohol measures**: CIDI at baseline. TLFB and short inventory of problems at followup. Potential subjects with drug use more severe than alcohol use disorder were excluded. Less than 10% HIV infected. Subjects lost to follow-up not included in the analysis.
Kalichman et al. 2008	**Population**: 342 men and women who drink in South African shebeens. **Setting**: Informal alcohol establishments (shebeens).	**Intervention**: 3-hour skills-based HIV–alcohol risk-reduction group session. **Control**: 1-hour HIV-alcohol information group session. **Assessment**: Baseline, 3, and 6 months.	The following behaviors were improved significantly at 3 months among the intervention group: • alcohol use before sex • unprotected intercourse • percent of sex with condoms • number of sex partners. Intervention effects were significantly stronger in those drinking less and dissipated at 6 months.	**Alcohol measures**: AUDIT, frequency of drinking before sex in previous month. Change in AUDIT scores not reported. 7% HIV infected in intervention group. 4% HIV infected in control group.

study had substantial methodological limitations, some of which pertain to the parent study. For example, in the parent study, random assignment of participants to the groups resulted in an imbalance between the groups with respect to baseline HIV risk behaviors or demographics. Moreover, the subanalysis was limited to participants who completed four follow-ups and were homeless or without stable housing. Finally, the outcome was alcohol or marijuana use in the last 3 months with no alcohol-specific results provided.

The four other studies in this category did not demonstrate any significant effects of the interventions tested on alcohol use:

- In a preliminary analysis of 3-month outcomes among 51 subjects randomized to four 1-hour motivational enhancement therapy sessions in an adolescent clinic, Naar-King and colleagues (2006) observed a trend, but no statistically significant reduction, in the number of drinks per week during the week with the

Table 2 Studies Identified but not Selected for the Literature Review

Citation	Population	Reason Excluded
Golin et al. 2003	140 HIV-infected patients. **Setting**: Hospital HIV clinic.	No data on the proportion of drinkers at baseline.
Goujard et al. 2003	326 HIV-infected patients. **Setting**: Hospital- and university-based centers.	No specific alcohol outcomes; alcohol group not analyzed independently.
Jones et al. 2003	174 women with AIDS from three U.S. cities recruited in 1997 from outpatient clinics, community health centers and agencies, and participant referrals. **Alcohol use**: 32% with history of alcohol. **Setting**: Primarily recruited from outpatient clinics, community health centers, and participant referrals.	No alcohol-specific outcomes reported.
Pradier et al. 2003	244 HAART-treated patients. **Setting**: Hospital	No specific alcohol outcomes; alcohol group not analyzed independently.
Samet et al. 2008	181 Russian men and women who reported any alcohol or drug dependence and who reported at least one incidence of unprotected sex in the past 6 months. **Setting**: Narcology hospitals	No alcohol-specific outcomes reported. Although both HIV-infected and alcohol-dependent patients were included in this study, the HIV-infected patients were not the alcohol-dependent patients.
Sampaio-Sa et al. 2008	107 HIV-infected, antiretroviral-naïve patients at an Brazilian HIV clinic for whom antiretrovirals were indicated were recruited from 2003 to 2004. 45% with alcohol use in the last 3 months.	Alcohol-specific outcomes not reported.
Simoni et al. 2007	136 HIV-infected men and women. **Setting:** Outpatient clinic	No information on current use; no specific alcohol outcomes.
Wong et al. 2008	936 HIV-infected from four U.S. cities recruited between 2000 and 2002. **Setting:** Community agencies, AIDS service organizations, and medical clinics	Alcohol-specific outcomes not reported; absolute numbers for outcome not presented.

SOURCES: Golin, C.E.; Earp, J.; Tien, H.C.; et al. A 2-arm, randomized, controlled trial of a motivational interviewing-based intervention to improve adherence to antiretroviral therapy (ART) among patients failing or initiating ART. *Journal of Acquired Immune Deficiency Syndromes* 42:42–51, 2006; Goujard, C.; Bernard, N.; Sohier, N.; et al. Impact of a patient education program on adherence to HIV medication: A randomized clinical trial. *Journal of Acquired Immune Deficiency Syndromes* 34:191–194, 2003; Jones, D.L.; Ishii, M.; LaPerriere, A.; et al. Influencing medication adherence among women with AIDS. *AIDS Care* 15:463–474, 2003; Pradier, C.; Bentz, L.; Spire, B.; et al. Efficacy of an educational and counseling intervention on adherence to highly active antiretroviral therapy: French prospective controlled study. *HIV Clinical Trials* 4:121–131, 2003; Samet, J.H.; Krupitsky, E.M.; Cheng, D.M.; 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* 103:1474–1483, 2008; Sampaio-Sa, M.; Page-Shafer, K.; Bangsberg, D.R.; et al. 100% adherence study: Educational workshops vs. video sessions to improve adherence among ART-naive patients in Salvador, Brazil. *AIDS and Behavior* 12:S54–S62, 2008; Simoni, J.M.; Pantalone, D.W.; Plummer, M.D.; and Huang, B. A randomized controlled trial of a peer support intervention targeting antiretroviral medication adherence and depressive symptomatology in HIV-positive men and women. *Health Psychology* 26:488–495, 2007; Wong, F.L.; Rotheram-Borus, M.J.; Lightfoot, M.; et al. Effects of behavioral intervention on substance use among people living with HIV: The Healthy Living Project randomized controlled study. *Addiction* 103:1206-1214, 2008.

maximum number of drinks. Moreover, in the final analysis of the study, which included 65 subjects, 39 percent of whom used alcohol, this difference was not sustained at 6 or 9 months (Naar-King et al. 2008).

- Gilbert and colleagues (2008) randomized 476 HIV-infected patients, 38 percent of whom reported risky drinking, to an MI-based "Video Doctor" intervention via laptop computer or a control group receiving usual care. The intervention resulted in decreased 30-day illicit drug use, lower mean number of drug use days, and a modest reduction of unprotected sex at 3 and 6 months. However, no differences in alcohol use existed between the intervention and control groups.

- Sorensen and colleagues (2003) randomly assigned HIV-infected patients with drug dependence, 61 percent of whom reported current alcohol use, to 1 year of continuous case management or to a brief contact (i.e., one HIV risk education session and printed information). No differences were noted in alcohol outcomes at 6, 12, or 18 months.

- A study among HIV-infected youths compared the effects of 23 2-hour group sessions and usual care on risk behaviors (Rotheram-Borus et al. 2001). The investigators found no changes from baseline on a measure reflecting alcohol and marijuana use and no difference between the intervention and control groups.

RCTs Among Alcohol Users at High-Risk for HIV Infection

Two informative RCTs have been conducted among alcohol drinkers at high risk for HIV infection. Morgenstern and colleagues (2007) performed a study with 198 high-risk, HIV-negative men who had sex with men and who were diagnosed with alcohol abuse or dependence but were seeking to moderate their alcohol use. The investigators compared the effects of 12 weekly MI sessions augmented with CBT with 4 sessions of MI alone. Unexpectedly, the investigators found that the nonaugmented MI group had less drinking and fewer alcohol-related drinking problems than the MI-plus-CBT group during the 12 weeks of the intervention and that there were no significant differences at 12-month follow-up. Thus, the addition of CBT to MI techniques provided no additional benefit regarding alcohol outcomes and potentially even diminished effects in this population. Subgroup analyses demonstrated that the detrimental effect of augmentation occurred particularly in participants with a concomitant drug use disorder.

Another RCT (Kalichman et al. 2008) compared a 3-hour, skills-based HIV and alcohol risk reduction group session with a 1-hour HIV/alcohol information group session among 342 South Africans frequenting drinking establishments. In this study, the extended session resulted in decreases in alcohol use before sex and unprotected intercourse at 3 month but not at 6 month follow-up. Moreover, intervention effects were stronger in participants drinking less at baseline.

Discussion

Given the high prevalence of unhealthy alcohol use among HIV-infected people and its associated adverse health consequences, development of clinical and public health interventions that seek to address alcohol use and improve health outcomes in this population is a priority. In recognition of this, NIAAA, as early as 1996, issued a request for applications entitled "Developing Alcohol-Related HIV Preventive Interventions (AA–97–03)." Since then, several studies have been published that describe clinical outcomes of interventions in this population. However, as this article has demonstrated, the literature on this important topic still is not extensive. A literature search revealed only four clinical intervention studies focusing exclusively on HIV-infected patients with current or past unhealthy alcohol use; five other clinical trials included and documented the alcohol use of some of their HIV-infected participants. Overall, the current state of research strongly suggests that although the problems related to alcohol in HIV-infected people are abundant, effective interventions are few and new ones are urgently needed. Hence, addressing alcohol problems remains an important issue in HIV research.

Not only are studies among alcohol-abusing, HIV-infected patients scarce, but the existing studies also yielded mixed results. Two of the four studies that specifically targeted HIV-infected people with alcohol problems showed improvement in drinking outcomes. Velasquez and colleagues (2009) demonstrated reduced drinking levels over 12 months after an intervention that included both MI and peer support. The intervention was particularly strong in reducing same-day drinking and sex, which compels further research on interventions targeting alcohol use at the time of HIV risk behaviors (Velasquez et al. 2009). Although the intervention types used in the study only were shown to be effective in a sample of men who have sex with men, they warrant study among other populations. In the other study, Aharanovich and colleagues (2006) demonstrated the feasibility of ongoing telephone-based interactive voice response and graphic feedback, which should inspire the inclusion of automated, tailored, ongoing intervention boosting as part of behavioral interventions. It is important to note, however, that both these studies had methodological limitations (e.g., substantial or differential loss to follow-up, incomplete assessments) and their findings therefore are not definitive. Nevertheless, they provide some guidance for future more rigorous clinical trials.

The other two clinical trials (Parsons et al. 2007; Samet et al. 2005) among alcohol-abusing HIV-infected people attempted to improve ART adherence. This is an appropriate target of alcohol

intervention studies in this population because medication adherence is of utmost importance for achieving good HIV disease outcomes, and alcohol-using patients have been documented to exhibit suboptimal ART adherence (Braithwaite et al. 2005; Chander et al. 2006; Conen et al. 2009; Samet et al. 2004). The results of both of these trials are discouraging, however, because although they explicitly addressed both alcohol use and medication adherence, one study (Samet et al. 2005) found no impact on adherence, alcohol consumption, or any HIV outcome, and the other (Parsons et al. 2007) only detected short-lived improvements (i.e., they were evident at 3 months, but not at 6 months). Thus, these two high-quality studies suggest that achieving clinically beneficial outcomes in HIV-infected people with alcohol problems is more difficult than has been the case with populations of HIV-infected without diagnosed unhealthy alcohol use (Amico et al. 2006; Simoni et al. 2006). Among the latter group, RCTs to improve adherence that used interventions with a range of intensities did reveal improvements in adherence which were sustained for up to 12 months, as well as in HIV viral load and CD4 counts (Tuldra et al. 2000). The difficulty of achieving positive benefits (e.g., improved ART adherence) through interventions among HIV-infected people who have alcohol problems also is evidenced by the study by Kalichman and colleagues (2008) among drinkers who were not infected with HIV. The findings of that study suggest that, as in brief intervention studies, intervention effectiveness varies by severity of alcohol use, with less improvement noted in dependent than in nondependent drinkers. Thus, levels of alcohol consumption, alcohol use disorder severity, and alcohol-related consequences are important covariates to be assessed and reported in HIV intervention studies.

A notable finding of this literature review was that as of 2009, no study of pharmacotherapy for alcohol dependence in HIV-infected patients had been published. This is surprising given that pharmacotherapy plays a major role in addressing the AIDS epidemic by improving outcomes of HIV-infected subjects. Moreover, some preclinical research has demonstrated that naltrexone, an effective medication for alcohol dependence, inhibits alcohol-mediated enhancement of HIV infection (Wang et al. 2006) and may potentiate the anti-HIV effects of antiretroviral medications (Gekker et al. 2001). Therefore, testing the effectiveness of naltrexone and other medications in alcohol-dependent HIV-infected patients is an important current research direction.

Two of the studies reviewed here that included HIV-infected patients among whom at least 10 percent currently used alcohol, targeted risky sexual behaviors rather than alcohol consumption. Assessing treatment effects on sex risk factors is appropriate for studies among HIV-infected drinkers because several studies have demonstrated an association between alcohol use and risky sex (Purcell et al. 2001; Stein et al. 2009). In both the study by Gilbert and colleagues (2008) and the study by Naar-King and colleagues (2006, 2008), sex risk behaviors were decreased in the group randomized to the intervention at 3 and 6 months, but there were no or only transient effects on alcohol use. These findings suggest that behavioral interventions which are not specifically tailored to address alcohol use are unlikely to impact alcohol problems in a sustained fashion.

The dearth of studies focusing on alcohol consumption among HIV-infected people is understandable. Although the spectrum of unhealthy alcohol use ranging from risky use to alcohol dependence occurs in this population, other pressing health concerns (e.g., ART adherence, risky sexual behaviors, or engagement in HIV care) appropriately become the main focus of clinical trials that also may address alcohol consumption in their intervention arms. Developing interventions that target a specific behavior (e.g., sex) at the time of alcohol use is a worthy pursuit, and understanding the importance of decreasing alcohol use in order to successfully achieve behavior change is crucial for developing future interventions.

One interesting development noted in the studies reviewed here was the use of new technology (e.g., interactive voice-response systems) in two of the studies (Aharonovich et al. 2006; Gilbert et al. 2008). These approaches to delivering a behavioral intervention merit further exploration because they have the potential for providing scalable, ongoing delivery of tailored automated messages that may boost a more intensive directly administered intervention.

When assessing the relevance of the studies reviewed here, particularly those conducted among HIV-infected patients with past or current unhealthy alcohol use, it is important to consider the methodological quality of the work (i.e., the potential for bias, design limitations, and outcome measures). The report by Velasquez and colleagues (2009) is the only controlled study demonstrating a sustained clinically significant treatment effect on an alcohol-specific outcome, making publication bias (i.e., the preferential publication of studies that find significant differences) unlikely.

Regarding their design, most, but not all, of these studies met important design criteria, such as random allocation of participants to treatment groups and intention-to-treat analyses[6] in the presentation of results. As with all behavioral intervention studies, keeping participants in the dark about which treatment they receive (i.e., blinding of participants to their treatment) is not possible. However, both Parsons and colleagues (2007) and Gilbert and colleagues (2008) utilized time- and content-equivalent controls to allow for the detection of effects

[6] An intention-to-treat analysis is based on the initial treatment intent, not on the treatment actually administered. Thus, every participant who begins the treatment is considered to be part of the trial, whether they finish it or not. This is done to avoid various misleading artifacts that can arise in a study. For example, if participants who have a more serious problem tend to drop out at a higher rate, even an ineffective treatment may appear to provide benefits if one only compares the condition before and after the treatment among participants who finish the treatment and ignores participants who were enrolled originally but did not finish the treatment.

specific to the counseling method studied.

The outcome measures reported were not consistent across studies and not always meaningful, limiting the comparability of study outcomes. For example, Naar-King and colleagues (2006) used an alcohol-specific measure—the number of drinks per week during the week with the maximum number of drinks at 3 months—that is not widely used and of questionable clinical meaning. Sorensen and colleagues (2003) only report a measure called the Addiction Severity Index Alcohol Composite Score, without any explanation or reporting of the individual components, complicating judgment of its clinical meaning. Finally, Samet and colleagues (2005) focused on ART adherence as an outcome, yet this study may underestimate the effectiveness of the intervention because the criteria for eligibility to participate in the study did not exclude patients with already good adherence. Thus, participants with good adherence at baseline provided little opportunity for an intervention to reveal a clinically meaningful impact.

In summary, as of 2009 the medical literature on clinical trials focused on people with HIV infection and unhealthy alcohol use is limited (i.e., "drops in a bottle"). Few of these studies were able to document improved outcomes, and any effects observed generally were modest and transitory. Based on these findings and current knowledge, the following questions need to be addressed:

- What are the characteristics of interventions that mitigate the health consequences of alcohol use in HIV-infected people?
- How does the treatment setting impact the effectiveness of behavioral interventions?
- How can technology best be used to extend and enhance intervention effects?
- What characteristics of HIV-infected drinkers suggest greater challenges when attempting to improve clinical outcomes?
- How can individual, network, or community interventions in people with multiple overlapping problems, including alcohol use, optimally reduce unhealthy behaviors?
- How might combined pharmacotherapy and behavioral therapy be utilized to address the spectrum of clinical consequences that accompany heavy alcohol consumption?

Obtaining answers to these questions is the key next step in the successful development of clinical and public health interventions to mitigate the adverse outcomes from alcohol use in HIV-infected patients. ■

Acknowledgements

The authors acknowledge Victoria Churchill, M.P.H, for her thoughtful assistance in the preparation of this manuscript and Carly Bridden, M.A., M.P.H. for reviewing the manuscript.

Financial Disclosure

The authors declare that they have no competing financial interests.

References

AHARONOVICH, E.; HATZENBUEHLER, M.L.; JOHNSTON, B.; ET AL. A low-cost, sustainable intervention for drinking reduction in the HIV primary care setting. *AIDS Care* 18:561–568, 2006. PMID: 16831783

AMICO, K.R.; HARMAN, J.J.; AND JOHNSON, B.T. Efficacy of antiretroviral therapy adherence interventions: A research synthesis of trials. 1996 to 2004. *Journal of Acquired Immune Deficiency Syndromes* 41:285–297, 2006. PMID: 16540929

ANTON, R.F.; O'MALLEY, S.S.; CIRAULO, D.A.; ET AL. Combined pharmacotherapies and behavioral interventions for alcohol dependence: The COMBINE study: A randomized controlled trial. *JAMA: Journal of the American Medical Association* 295:2003–2017, 2006. PMID: 16670409

BABOR, T.F.; HIGGINS-BIDDLE, J.C.; SAUNDERS, J.B.; AND MONTEIRO, M.G. *AUDIT the Alcohol Use Disorders Identification Test: Guidelines for use in Primary Care*. 2nd ed. Geneva: World Health Organization, 2001.

BEICH, A.; THORSEN, T.; AND ROLLNICK, S. Screening in brief intervention trials targeting excessive drinkers in general practice: Systematic review and meta-analysis. *BMJ* 327:536–542, 2003. PMID: 12958114

BRAITHWAITE, R.S.; MCGINNIS, K.A.; CONIGLIARO, J.; ET AL. A temporal and dose-response association between alcohol consumption and medication adherence among veterans in care. *Alcoholism: Clinical and Experimental Research* 29:1190–1197, 2005. PMID: 16046874

CHANDER, G.; LAU, B.; AND MOORE, R.D. Hazardous alcohol use: A risk factor for non-adherence and lack of suppression in HIV infection. *Journal of Acquired Immune Deficiency Syndromes* 43:411–417, 2006. PMID: 17099312

CONEN, A.; FEHR, J.; GLASS, T.R.; ET AL. Self-reported alcohol consumption and its association with adherence and outcome of antiretroviral therapy in the Swiss HIV Cohort Study. *Antiviral Therapy* 14:349–357, 2009. PMID: 19474469

CONIGLIARO, J.; GORDON, A.J.; MCGINNIS, K.A.; ET AL. 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* 33:521–525, 2003. PMID: 12869842

COOK, R.L.; SEREIKA, S.M.; HUNT, S.C.; ET AL. Problem drinking and medication adherence among persons with HIV infection. *Journal of General Internal Medicine* 16:83–88, 2001. PMID: 11251758

DAEPPEN, J.B.; GAUME, J.; BADY, P.; ET AL. Brief alcohol intervention and alcohol assessment do not influence alcohol use in injured patients treated in the emergency department: A randomized controlled clinical trial. *Addiction* 102:1224–1233, 2007. PMID: 17565563

D'ONOFRIO, G., AND DEGUTIS, L.C. Preventive care in the emergency department: Screening and brief intervention for alcohol problems in the emergency department: A systematic review. *Academic Emergency Medicine* 9:627–638, 2002. PMID: 12045080

EMMEN, M.J.; SCHIPPERS, G.M.; BLEIJENBERG, G.; AND WOLLERSHEIM, H. Effectiveness of opportunistic brief interventions for problem drinking in a general hospital setting: Systematic review. *BMJ* 328:318–322, 2004. PMID: 14729657

FLEMING, M.F.; BARRY, K.L.; MANWELL, L.B.; ET AL. Brief physician advice for problem alcohol drinkers: A randomized controlled trial in community-based primary care practices. *JAMA: Journal of the American Medical Association* 277:1039–1045, 1997. PMID: 9091691

GALVAN, F.H.; BING, E.G.; FLEISHMAN, J.A.; ET AL. 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, 2002. PMID: 12033694

GARBUTT, J.C.; KRANZLER, H.R.; O'MALLEY, S.S.; ET AL. Efficacy and tolerability of long-acting injectable naltrexone for alcohol dependence: A randomized controlled trial. *JAMA: Journal of the American Medical Association* 293:1617–1625, 2005. PMID: 15811981

GEKKER, G.; LOKENSGARD, J.R.; AND PETERSON, P.K. Naltrexone potentiates anti-HIV-1 activity of antiretroviral drugs in CD4+ lymphocyte cultures. *Drug and Alcohol Dependence* 64:257–263, 2001. PMID: 11672940

GILBERT, P.; CICCARONE, D.; GANSKY, S.A.; ET AL. Interactive "Video Doctor" counseling reduces drug and sexual risk behaviors among HIV-positive patients in diverse outpatient settings. *PLoS One* 3:e1988, 2008. PMID: 18431475

GOLIN, C.E.; LIU, H.; HAYS, R.D.; ET AL. A prospective study of predictors of adherence to combination antiretroviral medication. *Journal of General Internal Medicine* 17:756–765, 2002. PMID: 12390551

HALKITIS, P.N.; PARSONS, J.T.; WOLITSKI, R.J.; AND REMIEN, R.H. Characteristics of HIV antiretroviral treatments, access and adherence in an ethnically diverse sample of men who have sex with men. *AIDS Care* 15:89–102, 2003. PMID: 12655837

Institute of Medicine. *Broadening the Base of Treatment for Alcohol Problems: Report of a Study by a Committee of the Institute of Medicine Division of Mental Health and Behavioral Medicine.* Washington, DC: National Academy Press, 1990.

KALICHMAN, S.C.; ROMPA, D.; AND CAGE, M. Sexually transmitted infections among HIV seropositive men and women. *Sexually Transmitted Infections* 76:350–354, 2000. PMID: 11141850

KALICHMAN, S.C.; SIMBAYI, L.C.; VERMAAK, R.; ET AL. Randomized trial of a community-based alcohol-related HIV risk-reduction intervention for men and women in Cape Town South Africa. *Annals of Behavioral Medicine* 36:270–279, 2008. PMID: 18836789

KALICHMAN, S.C.; WEINHARDT, L.; DIFONZO, K.; ET AL. Sensation seeking and alcohol use as markers of sexual transmission risk behavior in HIV-positive men. *Annals of Behavioral Medicine* 24:229–235, 2002. PMID: 12173680

KANER, E.F.; BEYER, F.; DICKINSON, H.O.; ET AL. Effectiveness of brief alcohol interventions in primary care populations. *Cochrane Database of Systematic Reviews* CD004148, 2007. PMID: 17443541

KRANZLER, H.R., AND VAN KIRK, J. Efficacy of naltrexone and acamprosate for alcoholism treatment: A meta-analysis. *Alcoholism: Clinical and Experimental Research* 25:1335–1341, 2001. PMID: 11584154

KRISTENSON, H.; OHLIN, H.; HULTEN-NOSSLIN, M.B.; ET AL. Identification and intervention of heavy drinking in middle-aged men: Results and follow-up of 24-60 months of long-term study with randomized controls. *Alcoholism: Clinical and Experimental Research* 7:203–209, 1983. PMID: 6135365

LEFEVRE, F.; O'LEARY, B.; MORAN, M.; ET AL. Alcohol consumption among HIV-infected patients. *Journal of General Internal Medicine* 10:458–460, 1995. PMID: 7472704

LONGABAUGH, R.; WOOLARD, R.E.; NIRENBERG, T.D.; ET AL. Evaluating the effects of a brief motivational intervention for injured drinkers in the emergency department. *Journal of Studies on Alcohol* 62:806–816, 2001. PMID: 11838918

METSCH, L.R.; BELL, C.; PEREYRA, M.; ET AL. Hospitalized HIV-infected patients in the era of highly active antiretroviral therapy. *American Journal of Public Health* 99:1045–1049, 2009. PMID: 19372520

MIGUEZ, M.J.; SHOR-POSNER, G.; MORALES, G.; ET AL. HIV treatment in drug abusers: Impact of alcohol use. *Addiction Biology* 8:33–37, 2003. PMID: 12745413

MILLER, W.R., AND ROLLNICK, S. *Motivational Interviewing: Preparing People to Change*. New York: Guilford Press, 2002.

MONTI, P.M.; COLBY, S.M.; BARNETT, N.P.; ET AL. Brief intervention for harm reduction with alcohol-positive older adolescents in a hospital emergency department. *Journal of Consulting and Clinical Psychology* 67:989–994, 1999. PMID: 10596521

MORGENSTERN, J.; IRWIN, T.W.; WAINBERG, M.L.; ET AL. A randomized controlled trial of goal choice interventions for alcohol use disorders among men who have sex with men. *Journal of Consulting and Clinical Psychology* 75:72–84, 2007. PMID: 17295566

NAAR-KING, S.; LAM, P.; WANG, B.; ET AL. Brief report: Maintenance of effects of motivational enhancement therapy to improve risk behaviors and HIV-related health in a randomized controlled trial of youth living with HIV. *Journal of Pediatric Psychology* 33:441–445, 2008. PMID: 17905800

NAAR-KING, S.; WRIGHT, K.; PARSONS, J.T.; ET AL. Healthy choices: Motivational enhancement therapy for health risk behaviors in HIV-positive youth. *AIDS Education and Prevention* 18:1–11, 2006. PMID: 16539572

National Institute on Alcohol Abuse and Alcoholism. *Helping Patients Who Drink Too Much: A Clinician's Guide: Updated 2005 Edition.* Bethesda, MD: National Institutes of Health, 2007.

OCKENE, J.K.; ADAMS, A.; HURLEY, T.G.; ET AL. Brief physician- and nurse practitioner-delivered counseling for high-risk drinkers: Does it work? *Archives of Internal Medicine* 159:2198–2205, 1999. PMID: 10527297

OLMSTED, C.L., AND KOCKLER, D.R. Topiramate for alcohol dependence. *Annals of Pharmacotherapy* 42:1475–1480, 2008. PMID: 18698008

PARSONS, J.T.; GOLUB, S.A.; ROSOF, E.; AND HOLDER, C. Motivational interviewing and cognitive-behavioral intervention to improve HIV medication adherence among hazardous drinkers: A randomized controlled trial. *Journal of Acquired Immune Deficiency Syndromes* 46:443–450, 2007. PMID: 18077833

PROCHASKA, J.O, AND DICLEMENTE, C.C. Transtheoretical therapy—Toward a more integrative model of change. *Psychotherapy: Theory, Research and Practice* 19:276-288, 1982.

Project Match Research Group. Matching alcoholism treatments to client heterogeneity: Project MATCH posttreatment drinking outcomes. *Journal of Studies on Alcohol* 58:7–29, 1997. PMID: 8979210

PURCELL, D.W.; PARSONS, J.T.; HALKITIS, P.N.; ET AL. Substance use and sexual transmission risk behavior of HIV-positive men who have sex with men. *Journal of Substance Abuse* 13:185–200, 2001. PMID: 11547619

ROTHERAM-BORUS, M.J.; DESMOND, K.; COMULADA, W.S.; ET AL. Reducing risky sexual behavior and substance use among currently and formerly homeless adults living with HIV. *American Journal of Public Health* 99:1100–1107, 2009. PMID: 18799777

ROTHERAM-BORUS, M.J.; LEE, M.B.; MURPHY, D.A.; ET AL. Efficacy of a preventive intervention for youths living with HIV. *American Journal of Public Health* 91:400–405, 2001. PMID: 11236404

RUBIO, G.; JIMENEZ-ARRIERO, M.A.; PONCE, G.; AND PALOMO, T. Naltrexone versus acamprosate: One year follow-up of alcohol dependence treatment. *Alcohol and Alcoholism* 36:419–425, 2001. PMID: 11524308

SAITZ, R. Clinical practice: Unhealthy alcohol use. *New England Journal of Medicine* 352:596–607, 2005. PMID: 15703424

SAITZ, R.; HORTON, N.J.; CHENG, D.M.; AND SAMET, J.H. Alcohol counseling reflects higher quality of primary care. *Journal of General Internal Medicine* 23:1482–1486, 2008. PMID: 18618204

SAITZ, R.; PALFAI, T.P.; CHENG, D.M.; ET AL. Brief intervention for medical inpatients with unhealthy alcohol use: A randomized, controlled trial. *Annals of Internal Medicine* 146:167–176, 2007. PMID: 17283347

SAMET, J.H.; CHENG, D.M.; LIBMAN, H.; ET AL. Alcohol consumption and HIV disease progression. *Journal of Acquired Immune Deficiency Syndromes* 46:194–199, 2007. PMID: 17667330

SAMET, J.H.; FREEDBERG, K.A.; STEIN, M.D.; ET AL. Trillion virion delay: Time from testing positive for HIV to presentation for primary care. *Archives of Internal Medicine* 158:734–740, 1998. PMID: 9554679

SAMET, J.H.; HORTON, N.J.; MELI, S.; ET AL. A randomized controlled trial to enhance antiretroviral therapy adherence in patients with a history of alcohol problems. *Antiviral Therapy* 10:83–93, 2005. PMID: 15751766

SAMET, J.H.; HORTON, N.J.; MELI, S.; ET AL. Alcohol consumption and antiretroviral adherence among HIV-infected persons with alcohol problems. *Alcoholism: Clinical and Experimental Research* 28:572–577, 2004. PMID: 15100608

SAMET, J.H.; HORTON, N.J.; TRAPHAGEN, E.T.; ET AL. Alcohol consumption and HIV disease progression: Are they related? *Alcoholism: Clinical and Experimental Research* 27:862–867, 2003*a*. PMID: 12766632

SAMET, J.H.; LARSON, M.J.; HORTON, N.J.; ET AL. Linking alcohol- and drug-dependent adults to primary medical care: A randomized controlled trial of a multi-disciplinary health intervention in a detoxification unit. *Addiction* 98:509–516, 2003*b*. PMID: 12653820

SAMET, J.H.; PHILLIPS, S.J.; HORTON, N.J.; ET AL. Detecting alcohol problems in HIV-infected patients: Use of the CAGE questionnaire. *AIDS Research and Human Retroviruses* 20:151–155, 2004. PMID: 15018702

SIMONI, J.M.; PEARSON, C.R.; PANTALONE, D.W.; ET AL. Efficacy of interventions in improving highly active antiretroviral therapy adherence and HIV-1 RNA viral load: A meta-analytic review of randomized controlled trials. *Journal of Acquired Immune Deficiency Syndromes* 43(Suppl):S23–S35, 2006. PMID: 17133201

SORENSEN, J.L.; DILLEY, J.; LONDON, J.; ET AL. Case management for substance abusers with HIV/AIDS: A randomized clinical trial. *American Journal of Drug and Alcohol Abuse* 29:133–150, 2003. PMID: 12731685

STEIN, M.D.; ANDERSON, B.J.; CAVINESS, C.M.; ET AL. Relationship of alcohol use and sexual risk taking among hazardously drinking incarcerated women: An event-level analysis. *Journal of Studies on Alcohol and Drugs* 70:508–515, 2009. PMID: 19515290

TULDRA, A.; FUMAZ, C.R.; FERRER, M.J.; ET AL. Prospective randomized two-arm controlled study to determine the efficacy of a specific intervention to improve long-term adherence to highly active antiretroviral therapy. *Journal of Acquired Immune Deficiency Syndromes* 25:221–228, 2000. PMID: 11115952

VELASQUEZ, M.M.; VON STERNBERG, K.; JOHNSON, D.H.; ET AL. Reducing sexual risk behaviors and alcohol use among HIV-positive men who have sex with men: A randomized clinical trial. *Journal of Consulting and Clinical Psychology* 77:657–667, 2009. PMID: 19634959

WANG, X.; DOUGLAS, S.D.; PENG, J.S.; ET AL. Naltrexone inhibits alcohol-mediated enhancement of HIV infection of T lymphocytes. *Journal of Leukocyte Biology* 79:1166–1172, 2006. PMID: 16574767

WHITLOCK, E.P.; POLEN, M.R.; GREEN, C.A.; ET AL. Behavioral counseling interventions in primary care to reduce risky/harmful alcohol use by adults: A summary of the evidence for the U.S. Preventive Services Task Force. *Annals of Internal Medicine* 140:557–568, 2004. PMID: 15068985

WONG, F.L.; ROTHERAM-BORUS, M.J.; LIGHTFOOT, M.; ET AL. Effects of behavioral intervention on substance use among people living with HIV: The Healthy Living Project randomized controlled study. *Addiction* 103:1206–1214, 2008. PMID: 18494840

WUTZKE, S.E.; CONIGRAVE, K.M.; SAUNDERS, J.B.; AND HALL, W.D. The long-term effectiveness of brief interventions for unsafe alcohol consumption: A 10-year follow-up. *Addiction* 97:665–675, 2002. PMID: 12084136

AIDS Behav (2010) 14:S74–S83
DOI 10.1007/s10461-010-9723-y

ORIGINAL PAPER

Alcohol Use and Sex Risk Behaviors Among HIV-Infected Female Sex Workers (FSWs) and HIV-Infected Male Clients of FSWs in India

Jeffrey H. Samet · Christine A. Pace · Debbie M. Cheng · Sharon Coleman · Carly Bridden · Manoj Pardesi · Niranjan Saggurti · Anita Raj

Published online: 11 June 2010

Abstract Unprotected heterosexual transactional sex plays a central role in the spread of HIV in India. Given alcohol's association with risky sex in other populations and alcohol's role in HIV disease progression, we investigated patterns of alcohol use in HIV-infected female sex workers (FSWs) and HIV-infected male clients of FSWs in Mumbai. Analyses identified factors associated with heavy alcohol use and evaluated the relationship between alcohol use and risky sex. We surveyed 211 female and 205 male individuals; 80/211 FSWs (38%) and 127/205 male clients (62%) drank alcohol in the last 30 days. Among females, 32 and 11% drank heavily and were alcohol-dependent, respectively; among males the respective proportions were 44 and 29%. Men's heavy alcohol use was significantly associated with inconsistent condom use over the last year (AOR 2.40, 95% CI 1.21–4.77, $P = 0.01$); a comparable association was not seen in women. These findings suggest a need to address alcohol use both to avoid the medical complications of its heavy use in this population and to mitigate inconsistent condom use, the latter issue possibly requiring gender specific approaches. Such efforts to reduce drinking will be an important dimension to secondary HIV prevention in India.

Keywords Alcohol · Transactional sex · HIV · Female sex workers

J. H. Samet (✉) · D. M. Cheng · C. Bridden · A. Raj
Department of Medicine, Section of General Internal Medicine, Clinical Addiction Research and Education (CARE) Unit, Boston University School of Medicine/Boston Medical Center, 801 Massachusetts Ave., 2nd Floor, Boston, MA 02118, USA
e-mail: jsamet@bu.edu

J. H. Samet · A. Raj
Department of Community Health Sciences, Boston University School of Public Health, Boston, MA, USA

C. A. Pace
Internal Medicine Residency Program, Brigham and Women's Hospital, Boston, MA, USA

D. M. Cheng
Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA

S. Coleman
Data Coordinating Center, Boston University School of Public Health, Boston, MA, USA

M. Pardesi
Network of Maharashtra by People Living with HIV & AIDS (NMP+), Pune, India

N. Saggurti
Population Council, New Delhi, India

Background

In India, unprotected heterosexual transactional sex is a major risk behavior fueling the HIV epidemic [1]. Among Indian female sex workers (FSWs), the prevalence of HIV infection is about 15 times that of the general population [2]. Identifying factors associated with on-going high-risk sexual behavior among HIV-infected persons involved in transactional sex, both FSWs and their clients, is therefore a crucial component of HIV prevention efforts.

Studies from other countries document associations between alcohol use and high-risk heterosexual behaviors among those with and without HIV infection [3–6]. In India, only a minority of the population drinks any alcohol (i.e., 2% of women and 32% of men), although it is becoming more common [7, 8]. A national survey, however, suggests that FSWs and their male clients are more

likely than other groups in India to drink [2]. Furthermore, men who drink alcohol when visiting Indian FSWs engage in riskier sexual behavior (e.g., unprotected anal sex) and are more likely to have HIV and other sexually transmitted infections (STIs) [9]. Heavy alcohol use among HIV-infected persons has also been associated with increased risk of HIV disease progression as measured by CD4 cell count decline and absence of viral suppression [10–12].

These data suggest that interventions to reduce alcohol use may be an important component of primary HIV prevention efforts in India, but it is not clear whether alcohol use is common among FSWs and clients who are *already* infected with HIV, or whether alcohol plays a role in high-risk behaviors in this group. Independent of any relationship with risky sex, the medical morbidity due to heavy alcohol use among HIV-infected persons is another aspect of potential concern related to this population. Hence in this study, we survey HIV-infected FSWs and HIV-infected male clients of FSWs in Mumbai, with an aim of characterizing alcohol consumption in these groups. Analyses identified factors associated with heavy drinking and evaluated the association between alcohol consumption and risky sexual practices.

Methods

Recruitment and Enrollment

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. Female participants ($n = 211$) were recruited from the ASHA Center, a community based organization in Mumbai, run by a group of FSWs who provide support and linkage to care for HIV-infected sex workers and HIV-infected clients. The ASHA Center is associated with the HIV Positive People's Network in Maharashtra (NMP+), an NGO committed to the treatment and care of HIV-infected people in India. Male participants ($n = 205$) were recruited from ASHA and two other sites affiliated with NMP+.

HIV-infected outreach workers at the respective agencies reviewed client lists and selected every fifth individual from the list to be contacted (by phone or in person) for study participation. 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. Upon reaching the site, an outreach worker introduced the participant to gender-concordant research staff trained in HIV survey research. Two men and two women were excluded from screening due to intoxication. Of the 244 women and 370 men screened, 216 women and 210 men met eligibility criteria: 18 years or older; HIV-infected; and reporting sex trade involvement in the past year (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 did not complete their interviews and were thus excluded from the analyses, providing the final sample size of 211 female and 205 male participants ($n = 416$).

Human Participants Protections

This study was conducted as a partnership among Boston Medical Center, Boston University, Population Council, and NMP+. Procedures for this study were reviewed and approved by the institutional review boards of Boston University Medical Campus, Network of Maharashtra by People Living with HIV/AIDS (NMP+) and the Indian Council of Medical Research.

Subject Assessment

Participants received a 45 min interviewer-administered survey in Hindi assessing demographics, alcohol use, sexual 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 with consultation with the US investigators. Participants were also provided with 100 rupees ($2.50) as compensation for their time in this study.

Demographic and Health Information

Demographic data were collected based on items modified or taken from the Demographic and Health Survey [8] and Population Council surveys [13] and included gender, age, level of education, income, religion, marital status, and number of children. The survey included questions on the use of antiretroviral medications, and health status was assessed using the Hindi language Short-Form 12 Health Survey Version 2.0 (SF-12) [14], whose summary measures include the Mental Component Score (MCS) and Physical Component Score (PCS) [15]. The PCS and MCS scores have a range from 0 to 100 and were designed to have a mean score of 50 and a standard deviation of 10 in a representative sample of the US population [16].

Alcohol Outcomes

The survey included questions on alcohol use in various contexts in the past 3 months and past year. Daily consumption in the prior 7 days was collected using a

validated calendar method, the Timeline Follow Back (TLFB) [17], and was categorized as heavy, moderate, or abstinent. The "heavy" category was derived from the National Institute on Alcohol Abuse and Alcoholism definition of amounts that risk consequences (>14 drinks per week or >4 drinks on a single occasion for men, and >7 per week or >3 on a single occasion, for women) [18]. Heavy alcohol use (yes vs. no) was the primary outcome for analyses evaluating predictors of heavy drinking. Alcohol dependence over the last 12 months was evaluated using the Composite International Diagnostic Interview (CIDI) [19]. The survey asked about alcohol use prior to sexual encounters (e.g., "In the past 3 months, on how many occasions did you have a drink containing alcohol prior to having sex with a paid sex worker?") [20].

Sexual Behavior Outcomes

The primary sexual behavior outcome was the number of unprotected transactional sex encounters in the past 3 months (e.g., "In the past 3 months, how many times did you have sex with a client(s)?" "Of the times you had sex with a client in the past 3 months, how many times did you use a condom?"). Secondarily, the survey asked about the proportion of transactional sex encounters when the respondent used a condom (e.g. "In the past year, how often did you use a condom with a paid female sex partner?" "In the past year, how often did you use a condom with a client(s)?") with the response options of "always," "usually," "sometimes," "rarely," or "never." The secondary outcome of interest, derived from this question, was thus inconsistent condom use (yes vs. no). For this outcome, all responses other than "always" were categorized as inconsistent use.

Data Management

Data quality and management involved immediate review by field staff of survey data following interviews to ensure accuracy and completion as well as same day review by the field coordinator. Data entry and verification of consistency and accuracy utilized SPSS 11.5 software (SPSS Inc., Champaign, IL). Double data entry of the survey forms was performed in India and all discrepancies were reviewed and reconciled. Data were transmitted via a secure internet data transfer site to the Boston University School of Public Health, where further data cleaning and analysis occurred using SAS version 9.1 (SAS Institute Inc., Cary, NC).

Data Analysis

Identification of Factors Associated with Heavy Drinking. Heavy drinking over the past week, as measured by the Timeline Follow Back survey questions, was defined as the primary outcome. We used an iterative model building procedure based on logistic regression models to identify factors associated with this outcome. The following factors of interest were evaluated: age, income, religion, education, marital status, and MCS and PCS from the SF-12 v2. Age and PCS were modeled using tertiles as the linearity assumption did not hold in the regression models. We first fit unadjusted logistic regression models for each factor of interest. Factors significantly associated with heavy drinking at a significance level of 0.15 were included together in a single multivariable model. Factors that were no longer significant at the 0.15 level in the multivariable model were removed one at a time. Finally, factors not significant in unadjusted analyses were included one at a time in the multivariable model to assess their significance in the presence of other variables. The final model was determined using this iterative approach. Prior to regression modeling, we assessed bivariate correlations between all independent variables and covariates. To avoid potential colinearity, no pair of variables with Spearman correlation coefficient greater than 0.40 was included in the same model. Although a significance level criterion of 0.15 was used for entry and retention in the model building process, a two-sided alpha level of 0.05 was used to test whether a factor was significantly associated with heavy drinking.

Association between Heavy Drinking and Risky Sexual Behavior. The distribution of number of unsafe sex acts in the past 3 months, a count variable, was skewed, with a considerable proportion of zeros and a long tail. Thus, the use of models assuming normality was not appropriate. Therefore, we used Poisson regression models with overdispersion to evaluate the association between heavy drinking and the number of unsafe sex acts in the past 3 months [21].

The Pearson chi-square correction was used to account for overdispersion in the data. The regression models controlled for the following potential confounders: age, income, religion, education, and marital status. The magnitude of association between heavy drinking and number of unsafe sex acts was quantified using an incidence rate ratio (IRR). The IRR is the ratio of number of unsafe sex acts for heavy drinkers compared to those who were not heavy drinkers, thus the null value of no association for the IRR is equal to 1. The binary outcome inconsistent condom use was analyzed using multiple logistic regression models and resulting odds ratios were reported. These analyses were also controlled for age, income, religion, education and marital status. All analyses were conducted separately by gender. Analyses were performed using SAS software (version 9.1; SAS Institute, Cary, NC).

Results

Participant Characteristics and Alcohol Use

The demographic characteristics of the participants are shown in Table 1. In addition, participants reported frequency and quantity of alcohol consumed over the last 30 days and the last week. In the last 30 days, 38% (80/211; 95% CI 32–45%) of women and 62% (127/205; 55–69%) of men drank on at least one occasion. Among all FSWs, 32% (95% CI 26–38%) met the study threshold for a "heavy" drinking in the past week and 11% (23/211; 95% CI 7–15%) met criteria for alcohol dependence. Among the 80 women who reported any drinking, 84% (67/80) met the study threshold for a "heavy" level of drinking; 29% (23/80) met criteria for alcohol dependence. Among the men, 44% (95% CI 37–51%) met criteria for heavy drinking in the past week and 29% (59/205; 95% CI 23–36%) met criteria for alcohol dependence. Among the 127 drinkers, 71% (90/127) met criteria for heavy drinking and 46% [59/127] met criteria for alcohol dependence. Seventeen percent of the women and 40% of the men said they "always" or "usually" drank before having sex with a client/FSW (Table 1).

Condom Use

The FSWs in this study had many more transactional sex encounters than the male clients, and condoms were used in a lower proportion of the women's encounters. Women reported having had an average of 127 sexual encounters with clients over the last 3 months, while the men had an average of three encounters with paid female partners; in this time period the average number of unprotected sexual encounters was 15 for women and 0.2 for men. When asked about transactional sex practices over the last year, 90% of women and 26% of men reported inconsistent condom use (Table 1).

Demographic and Health Characteristics Associated with Alcohol Use

For the female sex workers, the final multivariable model examining predictors of heavy alcohol use revealed that younger age (AOR 4.69 for youngest tertile [22–28 years] vs. oldest tertile [32–50 years], 95% CI 2.14–10.30; OR 1.62 for middle tertile [29–31 years] versus oldest tertile [32–50 years], 95% CI 0.72–3.64, overall $P < 0.01$) and better self-reported physical health as measured by PCS (AOR 2.43 for highest tertile [42.6–54] versus lowest tertile [24.2–38.4], 95% CI 1.09–5.45; AOR 2.53 for middle tertile [38.5–42.5] versus lowest tertile [24.2–38.4], 95% CI 1.14–5.60, $P = 0.04$) were significantly associated with heavy alcohol use (Table 2).

For the male clients, none of the demographic or health measures were significantly associated with heavy alcohol use in the final multivariable model, though there was a trend toward a relationship between a lower income (i.e., less than the median for the group) and heavy alcohol use (AOR 1.68, 95% CI 0.95–2.98, $P = 0.07$) (Table 2).

Relationship Between Alcohol Use and Risky Sexual Behavior

Among the FSWs, no significant association was observed between those with heavy alcohol use and the reported number of unsafe transactional sex encounters over the past 3 months. The median (25th, 75th percentiles) number of such encounters was 0 (0, 20) for the heavy drinkers and 0 (0–11) for those with no heavy alcohol use ($P = 0.37$). The percentage who reported inconsistent condom use over the last year was also similar in both groups (90% of heavy drinkers and 91% of those who drank less or not at all). In the multivariable Poisson regression model (Table 3) heavy alcohol use in the past week was not associated with more unprotected transactional sex acts compared to those without heavy drinking ($P = 0.58$). Among other covariates in the model, younger age was significantly associated with the number of unprotected transactional sex encounters (IRR 2.94 for youngest tertile [22–28 years] vs oldest tertile [32–50 years], 95% CI 1.37–6.37; IRR 1.43 for middle tertile [29–31 years] vs oldest tertile [32–50 years], 95% CI 0.61–3.34; $P = 0.01$). We also did not detect an association between heavy alcohol use and the secondary outcome of inconsistent condom use in the past year ($P = 0.59$). Based on the multivariable logistic regression model, being currently unmarried was the only variable associated with inconsistent condom use (AOR 4.59, 95% CI 1.20–17.59, $P = 0.03$) (Table 4).

Heavy alcohol use and age played different roles in the analytic models for the male clients. Among the male clients, those with heavy alcohol use reported more unprotected transactional sex encounters over the last 3 months (IRR 2.75 [0.93–8.19], $P = 0.05$), an association that was borderline statistically significant. In addition, being in the youngest tertile (20–29 years) was associated with fewer unprotected sexual encounters over the last 3 months (IRR 0.14 for youngest tertile [20–29 years] vs. oldest tertile [36–49 years], 95% CI 0.02–0.95; IRR 0.04 for middle tertile [30–35 years] vs. oldest tertile [36–49 years], 95% CI 0.003–0.48, $P < 0.001$) (Table 3). For the secondary outcome inconsistent condom use (Table 4), both heavy drinking (AOR 2.40, 95% CI 1.21–4.77, $P = 0.01$) and older age (for youngest

Table 1 Characteristics of HIV-infected FSWs and HIV-infected male clients in India

	N (%)[a]	
	Female sex workers ($n = 211$)	Male clients ($n = 205$)
Demographic characteristics		
Age		
Mean (SD)	31 (5)	33 (6)
Religion		
Hindu	164 (78)	157 (77)
Muslim	25 (12)	19 (9)
Other	22 (10)	29 (14)
Currently married	20 (9)	76 (37)
Ever had children	83 (39)	100 (49)
Any formal schooling	46 (22)	182 (89)
Caste		
Scheduled caste	29 (14)	57 (28)
Scheduled tribe	24 (11)	24 (12)
Other backward caste	103 (49)	53 (26)
None	55 (26)	71 (34)
Monthly income (rupees/month)		
Median (range)	3000 (300–10000)	4500 (1500–20000)
Last CD4 count[b]		
Mean (SD)	384 (273)	294 (161)
Currently on antiretroviral medications	15 (7)	97 (47)
Mental health score[c]		
Mean (SD)	35 (6.8)	47 (5.9)
Physical health score[c]		
Mean (SD)	40 (5)	37 (3.4)
Alcohol use		
Any alcohol use in last 30 days[d]	80 (38)	127 (62)
Heavy alcohol use[e]	67 (32)	90 (44)
Alcohol dependence[f]	23 (11)	59 (29)
Drank before transactional sex encounters in last year		
Always/Usually	36 (17)	81 (40)
Sometimes/Rarely	52 (25)	56 (27)
Never	122 (58)	67 (33)
Sexual behavior		
No sexual encounters, past 3 months	3 (1.4%)	6 (2.9%)
Did not always use condoms in transactional sex encounters in last year	189 (90)	53 (26)
Number of unprotected transactional sex encounters in last 3 months		
Median (range)	0 (0–258)	0 (0–8)
Mean	14.9	0.14

[a] Except where means and medians are given, as noted

[b] Of 171 female and 190 male participants who knew their CD4 count

[c] Represent the MCS and PCS, respectively, from the SF-12 v2. Score range is 0–100; 50 is the mean score for both, with 0–49 below average and 51–100 above average

[d] The number who said they had had any alcohol in the last 30 days

[e] Participants were asked how much they drank in each of the last 7 days. Hazardous drinking is defined as >3 drinks in a day or >7 drinks/week for women and >4 drinks in a day or >14 drinks/week for men

[f] Based on the CIDI-SF, which asks about drinking and behaviors over the last year

tertile [20–29 years] vs. oldest tertile [36–49 years], AOR 0.46, 95% CI 0.19–1.11; for the middle tertile [30–35 years] versus oldest tertile [36–49 years], AOR 0.24, 95% CI 0.11–0.54; $P < 0.01$) were significantly associated with inconsistent condom use.

Discussion

Based on this survey examining key groups in the Indian AIDS epidemic, alcohol use is widespread among HIV-infected men and women involved in transactional sex in

Table 2 Characteristics associated with heavy alcohol use among HIV-infected FSWs and HIV-infected male clients in multivariable analyses

Characteristic	Female sex workers		Male clients	
	Adjusted odds ratio (95% CI)	*P* value	Adjusted odds ratio (95% CI)	*P* value
Age				
Youngest tertile[a]	4.69 (2.14–10.30)	<0.01	0.56 (0.27–1.17)	0.29
Middle tertile	1.62 (0.72–3.64)		0.82 (0.42–1.60)	
Oldest tertile	Referent		Referent	
Income				
<median number of rupees/month	N/A[b]	N/A	1.68 (0.95–2.98)	0.07
>/=median rupee/month	N/A		Referent	
Education				
No formal education	1.91 (0.85–4.26)	0.12	N/A	N/A
Formal education	Referent		N/A	
Physical health score				
Highest tertile[c]	2.43 (1.09–5.45)	0.04	N/A	N/A
Middle tertile	2.53 (1.14–5.60)		N/A	
Lowest tertile	Referent		N/A	

[a] Age tertiles (in years) for women: 22–28, 29–31, and 32–50; for men: 20–29, 30–35, and 36–49

[b] Items marked "N/A" did not meet criteria for inclusion into the multivariate model (see "Methods")

[c] Physical health score tertiles for women: 24.2–38.4, 38.5–42.5, and 42.6–54.0

Table 3 Characteristics associated with the number of unprotected transactional sex acts in the last 90 days among HIV-infected

Characteristic	Female sex workers		Male clients	
	Adjusted IRR[a] (95% CI)	*P* value	Adjusted IRR[a] (95% CI)	*P* value
Heavy alcohol use in last 7 days	0.83 (0.44–1.59)	0.58	2.75 (0.93–8.19)	0.05
Age				
Youngest tertile[b]	2.94 (1.37–6.37)	0.01	0.14 (0.02–0.95)	<.001
Middle tertile	1.43 (0.61–3.34)		0.04 (.003–0.48)	
Oldest tertile	Referent		Referent	
Income				
<median rupee/month	1.28 (0.67–2.42)	0.33	1.72 (0.63–4.68)	0.28
>/= median rupee/month	Referent		Referent	
Education				
No formal education	0.93 (0.47–1.83)	0.83	1.15 (0.24–5.56)	0.87
Formal education	Referent		Referent	
Marital status				
Currently unmarried	0.92 (0.34–2.50)	0.87	0.40 (0.13–1.20)	0.08
Currently married	Referent		Referent	
Religion				
Muslim and other	1.71 (0.89–3.26)	0.12	0.89 (0.22–3.60)	0.86
Hindu	Referent		Referent	

[a] IRR indicates incident rate ratio for number of unsafe paid sexual acts over last 90 days from multivariable Poisson regression model

[b] Age tertiles for women: 22–28, 29–31, and 32–50 years; tertiles for men: 20–29, 30–35, and 36–49

this country, and this use is commonly excessive. These findings sharply contrast with societal norms in which alcohol use is relatively uncommon in India as a whole, particularly among women. According to the 2005–2006 National Family Health Survey, which assesses behavioral risks in a nationally representative sample of adults in India, about 2% of women and 32% of men in India drink any alcohol [8]. The current data, however, are consistent with earlier data on the broader population of Indians involved in transactional sex, who have been found to be

Table 4 Characteristics associated with the outcome of inconsistent condom use over the last year among HIV-infected

Characteristic	Female sex workers		Male clients	
	Adjusted odds ratio (95% CI)	*P* value	Adjusted odds ratio (95% CI)	*P* value
Heavy alcohol use in last 7 days	0.75 (0.26–2.19)	0.59	2.40 (1.21–4.77)	0.01
Age				
Youngest tertile[a]	2.27 (0.69–7.51)	0.17	0.46 (0.19–1.11)	0.003
Middle tertile	2.97 (0.85–10.41)		0.24 (0.11–0.54)	
Oldest tertile	Referent		Referent	
Income				
< median rupee/month	0.35 (0.10–1.17)	0.09	0.76 (0.38–1.51)	0.43
>/= median rupee/month	Referent		Referent	
Education				
No formal education	1.82 (0.57–5.82)	0.31	2.17 (0.81–5.79)	0.12
Formal education	Referent		Referent	
Marital status				
Currently unmarried	4.59 (1.20–17.59)	0.03	0.77 (0.36–1.65)	0.50
Currently married	Referent		Referent	
Religion				
Muslim and others	0.96 (0.31–2.99)	0.94	0.55 (0.22–1.34)	0.19
Hindu	Referent		Referent	

[a] Age tertiles for women: 22–28, 29–31, and 32–50 years; tertiles for men: 20–29, 30–35, and 36–49

more likely to drink than the average Indian. The largest study to report such data, the National Behavioral Surveillance Study 2006, interviewed 7417 FSWs and 6613 male clients of FSWs without regard to HIV status and found that 46% of FSWs surveyed across India, and one-fourth of FSWs surveyed in the state of Maharashtra (which contains Mumbai) drank alcohol, while three-quarters of male clients of FSWs in India and two-thirds of male clients in Maharashtra drank alcohol [2]. That the prevalence of alcohol consumption in the current study of HIV-infected FSWs and their HIV-infected male clients was similar to this national cohort (in which presumably the majority were not infected with HIV) is consistent with research from other countries that has shown that heavy alcohol use is common among those infected with HIV [22–24].

In addition, a high percentage of participants—11% of the women and almost one-third of the men—were alcohol-dependent. About two-thirds of American adults drink alcohol, up to one-third of them at heavy levels, but only 4% have alcohol dependence [25–27].

It appears that despite presumably being in poorer health, no detectable decrease in percentage of persons drinking alcohol is observed among HIV-infected FSWs or their HIV-infected male partner compared to the findings in the FSWs and clients alcohol consumption survey.

For men and women involved in transactional sex, alcohol may play a variety of roles that could explain its common use. In particular, female sex workers working in brothels in Mumbai often offer alcohol to their male clients [Saggurti N.[Personal communication]. 6 Oct 2009] and indeed in this study most of the men who drink and a significant proportion of the women reported doing so in the setting of transactional sex. The logistic regression analysis was performed to help illuminate other factors that may play a role in heavy alcohol use and to reveal possible avenues to pursue in order to mitigate such consumption. Among the women, those who were younger and those who reported greater physical health (represented by higher PCS scores) had a higher odds of heavy drinking. This latter finding may reflect the "sick quitter" phenomenon in which less alcohol use occurs as one becomes too sick to drink, and thus overall those still drinking appear to reflect better health [28].

Although no associations were statistically significant, a notable magnitude of association was observed between lower income and heavy alcohol use among the men. Other studies have found that depression, family history, and personality traits (e.g., antisocial behavior) are related to heavy drinking among women, factors that we did not assess [29–32].

We did not detect an association between heavy drinking and unprotected sex among the FSWs; this finding may have a number of potential explanations. A recent meta-analysis of 27 studies of people with HIV in North America, Africa, Europe, and Russia found there to be a significant association between alcohol consumption (at all levels) and unprotected sex across the examined data sets

[5]. However, it is notable that the association in this meta-analysis was strongest in samples of men only. In the Indian population we studied, one explanation for the difference between the men and the women could be that male clients, rather than the sex workers, may more often control the final decision about condom use in sexual encounters, even if, as has been suggested in another study, the FSW may be more likely to propose condom use [33].

If men do control the final decision, condom use may be less likely in situations when men drink heavily and become disinhibited than in situations in which only the women drink. Alternatively, or in addition, it may be that the study's methodology affected its ability to discern a relationship between heavy drinking and unprotected sex among the female sex workers. In particular, it is important to note that this analysis looked *globally* at the relationship between alcohol intake and inconsistent condom use and, given the high proportion of women who did not consistently have safe sex (90%), detecting an association between the two behaviors would thus be difficult. It is possible that event-level questions asking about individual sexual encounters might have allowed detection of an association between the two behaviors. However, similar questions to those in this study have detected associations between alcohol and unprotected sex in other settings, so the question format alone is unlikely to account entirely for the lack of association among the women [5]. Also of note, this analysis looked only at a self-reported high-risk sexual behavior (i.e., condom use); it is possible that biologic markers such as sexually transmitted infections, might reveal an association between alcohol and risky sex [34].

Beyond the issue of the association between alcohol consumption and increased risk of HIV transmission among HIV-infected men and women involved in transactional sex in Mumbai, it is important to remain vigilant about the fact that high levels of alcohol use have important health implications for HIV-infected persons. Heavy drinking is known to confer high risk for a spectrum of psychological, social, and health problems [27], including adverse HIV-related health behaviors such as poor antiretroviral medication adherence [35, 36], and poor attendance at medical appointments [37, 38]. Heavy alcohol use among HIV-infected persons has also been associated with increased risk of medical problems including worse depressive symptoms [39] and poorer HIV immunological function (i.e., CD4 cell count decline) [10, 11]. Thus, in addition to alcohol's effects on sexual behavior, the heavy alcohol use found so commonly among the drinkers in this study may negatively affect their HIV-related health outcomes, as well as viral suppression and CD4 cell count. Future research could help define these implications in this population and illuminate the potential benefits of an intervention to reduce alcohol use.

Findings from this study demonstrate a need for programs for those involved in sex work that address alcohol consumption and condom use. The findings also suggest that within care systems for HIV-infected Indians, providers should screen for alcohol problems with subsequent brief interventions for risky drinkers and link those with alcohol dependence to alcohol treatment. Venue based interventions (i.e., interventions in alcohol or sex work settings) may also be useful to address risky alcohol behaviors for both FSWs and male clients, regardless of HIV status, but alone these could be inadequate, as the majority of men are likely drinking prior to entry into the venue. Hence, among men, social norm campaigns are needed to encourage condom use even in social situations in which both alcohol use and transactional sex occurs. FSWs, in contrast, may be using alcohol less for social reasons and more as a means of coping with difficult work situations, as noted in other studies [40–42]. Thus, interventions to reduce FSWs' use of alcohol would require approaches that alter the working conditions for this population. Nonetheless, it is unclear if such an approach would additionally impact HIV risk, as current findings revealed no association between alcohol and condom use among FSWs.

There are several limitations to this study. First, as indicated above, it relied on self-reporting of both alcohol and sexual risk behaviors. Under-reporting of stigmatized behaviors, resulting in an underestimate of the prevalence of alcohol use and condom use, could lead to an under-appreciation of an association between heavy drinking and unprotected sex. By using trained interviewers rather than clinical staff, we sought to minimize this phenomenon. Another limitation is that although when possible we used standard research tools and instruments previously used in India, the survey did not exclusively include instruments validated in Hindi. A third limitation is that the study surveyed a group of HIV-infected men and women who were members of an HIV service network. They therefore may have been more organized or felt more responsible for their health than other HIV-infected individuals involved in transactional sex. In addition, these study participants had already received some HIV prevention education and health care services, which may have reduced their risky sexual behavior and potentially the extent of their alcohol use. If these possibilities were at play during the conduct of this study, the result again would be an underestimate, rather than an overestimate, of the alcohol use and risky sexual behavior among this population. It is noteworthy that even if this is a "best case scenario," much remains to be done to address excessive alcohol use and unsafe sex in these individuals who are at high risk for transmitting HIV infection. A fourth limitation is that the major outcomes for alcohol consumption and condom use had different

time-frames. The study used "alcohol consumption over the last week" to define heavy drinking, whereas the two questions about condom use asked about behavior over the last 3 months and last year, respectively. Thus, the analysis of the relationship between heavy drinking and risky sexual practices relies on the assumption that participants' drinking over the last week also reflects the alcohol consumption patterns over the last 3 months and past year. A final limitation, as noted above, is the use of global-level questions about alcohol and sexual practices in the survey; it is possible that more detailed examination of situation-level or event-level behaviors (e.g., of concurrent drinking and unprotected sex) might have revealed a stronger association between alcohol and high-risk sexual behavior.

In summary, the current study makes a novel contribution by providing a detailed characterization of participants' alcohol use, revealing a strikingly high prevalence of heavy drinking among HIV-infected female sex workers and their HIV-infected male clients in Mumbai, India, despite low levels of drinking nationwide. Heavy drinking by these men appears to be associated with lower rates of condom use. These findings raise the possibility that efforts to reduce alcohol use among HIV-infected men who purchase heterosexual sex in India may be an important component of secondary HIV prevention initiatives. In addition, given the known effects of heavy alcohol consumption on HIV-related health outcomes, interventions to reduce drinking among HIV-infected men and women involved in transactional sex may improve the health of this vulnerable group in India.

Acknowledgments Funded by the National Institute on Alcohol Abuse and Alcoholism (R01AA016059-S1 and K24AA015674) and the Indian Council of Medical Research. We would like to thank Ms. Shabana Patel from the Network in Maharashtra by People Living with HIV (NMP+) for her oversight of study implementation in terms of helping ensure safe, confidential, and respectful treatment of our HIV-infected participants. We would like to thank the members of our partnering community based organizations in Mumbai for their assistance in implementing this study in the field, including the recruitment sites—ASHA, District Level Network registry, Drop-in HIV Center at the King Edward Memorial Hospital. 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. We thank Ms. Victoria Churchill and Mr. Joseph Novotny for their assistance with manuscript preparation. Finally, we are immensely grateful to the participants of this study who have, without hesitation, given their time and narrated their experiences and views regarding alcohol use and sexual behaviors.

References

1. UNAIDS. Report on the global AIDS epidemic. 2008. http://www.unaids.org/en/KnowledgeCentre/HIVData/GlobalReport/2008/2008_Global_report.asp. Accessed 15 Dec 2009.
2. National AIDS Control Organisation. National Behavioural Surveillance Survey (BSS): Female Sex Workers (FSWs) and their Clients. 2006. http://www.nacoonline.org/upload/NACO%20PDF/Female_Sex_Workers_%28FSWs%29_and_Their_Clients.pdf. Accessed 13 Aug 2009.
3. Weiser SD, Leiter K, Heisler M, et al. A population-based study on alcohol and high-risk sexual behaviors in Botswana. PLoS Med. 2006;3(10):e392.
4. Kiene SM, Simbayi LC, Abrams A, Cloete A, Tennen H, Fisher JD. High rates of unprotected sex occurring among HIV-positive individuals in a daily diary study in South Africa: the role of alcohol use. J Acquir Immune Defic Syndr. 2008;49(2):219–26.
5. Shuper P, Joharchi N, Irving H, Rehm J. Alcohol as a correlate of unprotected sexual behavior among people living with HIV/AIDS: review and meta-analysis. AIDS Behav. 2009;13(6):1021–36.
6. Ehrenstein V, Horton NJ, Samet JH. Inconsistent condom use among HIV-infected patients with alcohol problems. Drug Alcohol Depend. 2004;73(2):159–66.
7. Prasad R. Alcohol use on the rise in India. Lancet. 2009;373:17–8.
8. International Institute for Population Sciences (IIPS) and Macro International. National Family Health Survey (NFHS-3): India. Mumbai: IHS. 2005–2006. http://www.nfhsindia.org/NFHS-3%20Data/VOL-1/India_volume_I_corrected_17oct08.pdf. Accessed 9 Mar 2009.
9. Madhivanan P, Hernandez A, Gogate A, et al. Alcohol use by men is a risk factor for the acquisition of sexually transmitted infections and human immunodeficiency virus from female sex workers in Mumbai, India. Sex Transm Dis. 2005;32(11):685–90.
10. 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):194–9.
11. Chander G, Lau B, Moore RD. Hazardous alcohol use: a risk factor for non-adherence and lack of suppression in HIV infection. J Acquir Immune Defic Syndr. 2006;43(4):411–7.
12. Braithwaite RS, Conigliaro J, Roberts MS, et al. Estimating the impact of alcohol consumption on survival for HIV+ individuals. AIDS Care. 2007;19:459–66.
13. Population Council. Assessing the patterns and drivers of migration/mobility of sex workers and male workers and examining the links with HIV risk. http://www.popcouncil.org/projects/HIV_IndiaMigrationHIV.html. Accessed 30 Dec 2009.
14. Ware JE 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(3):220–33.
15. Ware JE Jr, Kosinski M, Keller SD. SF-36 physical and mental health summary scales: a user's manual. Boston, MA: The Health Inst; 1994.
16. Ware JE Jr, Gandek B, Kosinski M, et al. The equivalence of SF-36 summary health scores estimated using standard and country-specific algorithms in 10 countries: results from the IQOLA Project. J Clin Epidemiol. 1998;51(11):1167–70.
17. Sobell LC, Sobell MB. Alcohol Timeline Followback (TLFB) users' manual. Toronto, Canada: Addiction Research Foundation; 1995.
18. National Institute on Alcohol Abuse and Alcoholism. Helping patients who drink too much: a clinician's guide: updated 2005 edition. Bethesda, MD: National Institutes of Health. 2007. http://pubs.niaaa.nih.gov/publications/Practitioner/CliniciansGuide2005/guide.pdf. Accessed 27 Oct 2009.
19. Robins LN, Wing J, Wittchen HU, et al. 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(12):1069–77.

20. Raj A, Reed E, Santana MC, et al. The associations of binge alcohol use with HIV/STI risk and diagnosis among heterosexual African American men. Drug Alcohol Depend. 2009;101:101–6.
21. McCullagh P, Nelder JA. Generalized linear models. 2nd ed. New York: Chapman and Hall/CRC; 1989.
22. Conigliaro J, Gordon AJ, McGinnis KA, Rabeneck L, Justice AC. Veterans Aging Cohort 3-Site Study. How harmful is hazardous alcohol use and abuse in HIV infection: do health care providers know who is at risk? J Acquir Immune Defic Syndr. 2003;33(4): 521–5.
23. Krupitsky EM, Horton NJ, Williams EC, et al. Alcohol use and HIV risk behaviors among HIV-infected hospitalized patients in St. Petersburg, Russia. Drug Alcohol Depend. 2005;79(2):251–6.
24. 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): 151–5.
25. National Institute on Alcohol Abuse and Alcoholism. National epidemiologic survey on alcohol and related conditions: alcohol alert. 2006. http://pubs.niaaa.nih.gov/publications/AA70/AA70.htm. Accessed 12 Oct 2009.
26. Grant BF, Dawson DA, Stinson FS, Chou SP, Dufour MC, Pickering RP. The 12-month prevalence and trends in DSM-IV alcohol abuse and dependence: United States, 1991–1992 and 2001–2002. Drug Alcohol Depend. 2004;74(3):223–34.
27. Saitz R. Clinical practice. Unhealthy alcohol use. N Engl J Med. 2005;352(6):596–607.
28. Shaper AG, Wannamethee G, Walker M. Alcohol and mortality in British men: explaining the U-shaped curve. Lancet. 1988;2(8623):1267–73.
29. Dixit AR, Crum RM. Prospective study of depression and the risk of heavy alcohol use in women. Am J Psychiatry. 2000;157: 751–8.
30. Wilsnack SC, Klassen AD, Schur BE, Wilsnack RW. Predicting onset and chronicity of women's problem drinking: a five-year longitudinal analysis. Am J Public Health. 1991;81:305–18.
31. Poikolainen K. Risk factors for alcohol dependence: a case-control study. Alcohol Alcohol. 2000;35:190–6.
32. National Institute on Alcohol Abuse and Alcoholism. The genetics of alcoholism: alcohol alert. 2003. http://pubs.niaaa.nih.gov/publications/aa60.htm. Accessed 20 May 2010.
33. Dandona R, Dandona L, Gutierrez JP, et al. High risk of HIV in non-brothel based female sex workers in India. BMC Public Health. 2005;5:87.
34. Ghebremichael M, Paintsil E, Larsen U. Alcohol abuse, sexual risk behaviors, and sexually transmitted infections in women in Moshi urban district, northern Tanzania. Sex Transm Dis. 2009;36:102–7.
35. 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):572–7.
36. Hendershot CS, Stoner SA, Pantalone DW, Simoni JM. Alcohol use and antiretroviral adherence: review and meta-analysis. J Acquir Immune Defic Syndr. 2009 [epub ahead of print].
37. Mugavero MJ, Lin HY, Willig JH, et al. Missed visits and mortality among patients establishing initial outpatient HIV treatment. Clin Infect Dis. 2009;48(2):248–56.
38. Cunningham WE, Sohler NL, Tobias C, et al. Health services utilization for people with HIV infection: comparison of a population targeted for outreach with the U.S. population in care. Med Care. 2006;44(11):1038–47.
39. Sullivan LE, Saitz R, Cheng DM, Libman H, Nunes D, Samet JH. The impact of alcohol use on depressive symptoms in human immunodeficiency virus-infected patients. Addiction. 2008; 103(9):1461–7.
40. Li Q, Li X, Stanton B. Alcohol use among female sex workers and male clients: an integrative review of global literature. Alcohol Alcohol. 2010;45:188–99.
41. World Health Organization. A rapid situation assessment of sexual risk behaviour and substance use among sex workers and their clients in Chennai (Madras), South India. http://www.who.int/mental_health/evidence/sexual_behaviour_assessment_chennai.pdf. Accessed 20 May 2010.
42. Gupta J, Raj A, Decker MR, Reed E, Silverman JG. HIV vulnerabilities of sex-trafficked Indian women and girls. Int J Gynaeco Obstet. 2009;107:30–4.

Pulmonary Vascular Resistance, Collateral Flow, and Ventricular Function in Patients With a Fontan Circulation at Rest and During Dobutamine Stress

Boris Schmitt, MD; Paul Steendijk, PhD; Stanislav Ovroutski, MD; Karsten Lunze, MD, MPH; Pedram Rahmanzadeh, MD; Nizar Maarouf, MD; Peter Ewert, MD; Felix Berger, MD, PhD; Titus Kuehne, MD, PhD

Background—The role, interplay, and relative importance of the multifactorial hemodynamic and myocardial mechanisms causing dysfunction of the Fontan circulation remain incompletely understood.

Methods and Results—Using an MRI catheterization technique, we performed a differential analysis of pulmonary vascular resistance and aortopulmonary collateral blood flow in conjunction with global ventricular pump function, myocontractility (end-systolic pressure-volume relation), and diastolic compliance (end-diastolic pressure-volume relation) in 10 patients with a Fontan circulation at rest and during dobutamine stress. Pulmonary and ventricular pressures were measured invasively and synchronized with velocity-encoded MRI-derived pulmonary and aortic blood flows and cine MRI-derived ventricular volumes. Pulmonary vascular resistance and end-systolic and end-diastolic pressure-volume relations were then determined. Aortopulmonary collateral flow was calculated as the difference between aortic and pulmonary flow. Compared to rest, dobutamine caused a small increase in mean pulmonary pressures ($P<0.05$). Collateral flow was significantly augmented ($P<0.001$) and contributed importantly to an increase in pulmonary flow ($P<0.01$). Pulmonary vascular resistance decreased significantly ($P<0.01$). Dobutamine did not increase stroke volumes significantly despite slightly enhanced contractility (end-systolic pressure-volume relation). Active early relaxation (τ) was inconspicuous, but the end-diastolic pressure-volume relation shifted upward, indicating reduced compliance.

Conclusions—In patients with a Fontan circulation, aortopulmonary collateral flow contributes substantially to enhanced pulmonary flow during stress. Our data indicate that pulmonary vascular response to augmented cardiac output was adequate, but decreased diastolic compliance was identified as an important component of ventricular dysfunction. **(*Circ Cardiovasc Imaging*. 2010;3:623-631.)**

Key Words: heart defects, congenital ■ pulmonary heart disease ■ heart failure ■ magnetic resonance imaging

Ventricular function and pulmonary vascular resistance (PVR) are important variables that determine the outcome of patients with a Fontan circulation. Impaired growth of the pulmonary arteries failing to match somatic growth was described in recent studies.[1,2] Other authors have reported significant left-to-right shunting through aortopulmonary collaterals.[3,4] Both factors may have a direct impact on the PVR that cannot be measured in the Fontan circulation by conventional techniques, such as thermodilution or oxymetry.

Clinical Perspective on p 631

In addition to pulmonary vascular factors, heart failure was identified as an important cause of morbidity, and the right-type systemic ventricle seems to carry a higher risk for developing heart failure than the left-type ventricle.[5,6] However, the onset, course, and predominant form of heart failure vary and remain unpredictable. Systolic dysfunction has been described, but the role and mechanisms of diastolic dysfunction are less known.[7–10] Diastolic filling is considered to be impaired due to low preload that by itself is at least partly controlled by PVR. In addition, ventricular compliance is likely to be altered because congenital abnormalities as well as prolonged cyanosis and volume load may have an impact on the fibrous matrix of the myocardium.[7,11]

In this study, we performed a differential analysis of PVR and aortopulmonary collateral blood flow in conjunction with global ventricular pump function, ventricular contractility, and diastolic compliance. Measurements were obtained at rest

Received December 15, 2009; accepted June 1, 2010.

From the Department of Congenital Heart Disease and Pediatric Cardiology (B.S., S.O., K.L., P.R., N.M., P.E., F.B., T.K.), Deutsches Herzzentrum Berlin, and The Charité, Medical University Berlin, Berlin, Germany, and Departments of Cardiology and Cardiothoracic Surgery (P.S.), Leiden University Medical Center, Leiden, The Netherlands.

Correspondence to Titus Kuehne, Unit of Cardiovascular Imaging, Department of Congenital Heart Disease and Pediatric Cardiology, Deutsches Herzzentrum Berlin, Augustenburger Platz 1, 13353 Berlin, Germany. E-mail kuehne@dhzb.de

Circ Cardiovasc Imaging **is available at http://circimaging.ahajournals.org** **DOI: 10.1161/CIRCIMAGING.109.931592**

and during dobutamine stress using an MRI catheterization technique. We hypothesized that exposure to dobutamine stress would have a specific impact on PVR and aortopulmonary collateral flow as well as on systolic and diastolic ventricular function and, hence, improve our understanding of cardiovascular pathophysiology in patients with a Fontan circulation.

Methods

Study Design

The study was conducted in 10 preselected patients with a Fontan circulation. The patients were experiencing decreasing exercise capacity at routine follow-up and, thus, were referred to our institution for cardiac catheterization and MRI to determine ventricular systolic and diastolic function and cardiovascular anatomy. The exercise capacity assessment was based on New York Heart Association classification and ergometry. Only patients with extracardiac total cavopulmonary connection that had no fenestration of the tunnel were included. Exclusion criteria were the presence of pulmonary artery stenosis, atrioventricular valve insufficiency, arrythmias, protein-loss syndrome, thromboembolism, effusion, or edema. In addition, patients on β-blocker medication were not included because β-blockers would have affected the impact of dobutamine infusion. All patients or their guardians gave informed consent for the study, which was approved by the responsible institutional review board and ethics committee.

Atrioventricular valvar function and ventricular inflow profiles (E/A-wave ratio) were evaluated by Doppler echocardiography. Exercise capacity was quantified by oxygen uptake during ergometry. During catheterization, ventricular pressures were measured with 4- to 5-F fluid-filled pigtail catheters and pulmonary pressures with 5-F wedge catheters. Catheters were advanced over a femoral artery and femoral or cubital vein approach, respectively. At the end of the study, a 24- to 36-mm sizing balloon was placed at the level of the tunnel-to-pulmonary artery connection. Patients then were transferred to the neighboring MRI laboratory, keeping all catheters in place. Details about the use of the catheters during MRI follow.

MRI Assessment of Blood Flow and PVR

All MRI studies were performed with a 1.5-T scanner, with a maximum gradient performance of 30 mT/m and slew rate of 150 T/m per second. A 5-element cardiac phased-array coil was used for signal acquisition. In all patients, MRI was performed without sedation. Quantitative blood flow was measured using velocity-encoded (VEC) MRI orthogonal to the dominating flow direction in the superior and inferior vena cave, left and right pulmonary artery, and the ascending aorta as reported elsewhere.[12–14] In the ascending aorta, flow was measured just distal to the coronary arteries and the semilunar valves. In the inferior and superior vena cava, flow was measured in a segment approximately 5 to 10 mm below or above their connection to the pulmonary artery. Pulmonary flow was measured simultaneously with invasive pulmonary pressure. Technical details for pressure recording follow. All measurements were performed during free breathing at rest and during continuous infusion of 10 μg/kg per minute dobutamine. Automated correction was performed for potential phase errors arising from the concomitant magnetic field. Sequence parameters were TR and TE shortest, acquired spatial resolution of 2×2×6 mm, 2 numbers of excitation, phase-encoding velocity of 70 cm/s for venous and pulmonary flow and 150 cm/s for aortic flow, and 35 reconstructed phases per cardiac cycle.

Data analysis was done with View Forum release 6.1 software. Antegrade and retrograde flows were measured as described elsewhere.[12,15]

Aortopulmonary collateral flow (mL) was defined as the difference between effective antegrade aortic flow and the sum of antegrade right and left pulmonary blood flow as measured with VEC MRI in the right and left pulmonary artery.

PVR was the quotient between transpulmonary gradient and effective antegrade pulmonary flow.[16,17] The transpulmonary gradient was calculated as the difference between mean pulmonary and ventricular end-diastolic pressures. Total effective pulmonary flow was defined as the sum of antegrade flow as measured with VEC MRI in the right and left pulmonary artery plus collateral blood flow.

The Nakata index was calculated as the sum of the cross-sectional areas of the left and right pulmonary artery divided by body surface area.[18] The areas were obtained from the magnitude images of the right and left pulmonary VEC MRI measurements.

MRI Assessment of Global Ventricular Function

Ventricular chamber volumes and myocardial mass were determined from axial stacks of multislice-multiphase steady-state free precession cine MRI covering the entire heart.[12,19,20] Sequence parameters were TR and TE, 3.4 and 1.7 milliseconds; slice thickness, 6 mm; no gap; in-plane resolution, 1.9×1.3 mm^2; 45 phases per cardiac cycle; number of averages, 1; and sensitivity encoding reduction factor, 2. Analysis was done using View Forum release 6.1 software. Biventricular endo- and epicardial borders were manually traced for computing ventricular volumes and myocardial mass, where the septum was accounted for as left ventricular mass. Papillary muscles and prominent right ventricular trabeculation were excluded for volume measurements. Stroke volume was calculated as the difference between the diastolic and systolic volumes. Ejection fraction was calculated as the ratio of stroke volume to end-diastolic volumes.

MRI Assessment of Myocontractile and Diastolic Function

Parameters of myocontractile and diastolic function were derived from pressure-volume loops that were measured by an MRI catheterization technique. This MRI method was previously validated in animal experiments and has been recently described in detail.[21] Briefly, ventricular volumes of the single ventricle were acquired over several cardiac beats with cine MRI. During MRI, invasive ventricular pressures were measured and averaged. At postprocessing, volumes and pressures were synchronized in time using a trigger signal. From these measures, a pressure-volume loop under steady-state conditions was constructed. The end-systolic pressure-volume relation was estimated from this loop using a single-beat approach as previously described.[22–24] Emax, which is the slope of the end-systolic pressure-volume relation, was defined as a measure of contractility and indexed to 100 mg myocardial muscle mass (Emax,i). Effective arterial elastance (Ea) was calculated as end-systolic pressure divided by stroke volume. Ventricular-arterial coupling was determined as the ratio of Emax to Ea.[23,25]

In a second step, instantaneous blood flows were measured using real-time VEC MRI in the ascending aorta just distal to the orifices of the coronary arteries.[13,21] The sequence parameters were TR and TE, 23 and 6.5 milliseconds; matrix, 128×256; field of view, 400 mm; slice thickness, 8 mm; encoding velocity, 150 cm/s; sensitivity encoding reduction factor, 3; half-scan factor, 0.6; and echo planar imaging factor, 41. The scan time was 31 milliseconds for the acquisition of 1 phase-contrast image. During flow measurements, ventricular preload was reduced by transient balloon occlusion of the vena cava. The balloon was inflated with isotonic saline solution. For each unloaded beat, we determined ventricular chamber volumes and synchronized them with ventricular pressures to generate a set of pressure-volume loops, as previously described.[21] The absolute end-diastolic volumes were determined by matching the VEC MRI-derived volumes with the intercept of the end-systolic pressure-volume relation. The resulting end-diastolic pressure-volume points were used to determine the end-diastolic pressure-volume relation. The stiffness constant ß, a load-independent measure of ventricular compliance, was calculated from the end-diastolic pressure-volume relation with an exponential regression as follows: $EDP = Ae^{ß \cdot EDV}$, where EDP indicates end-diastolic pressure; EDV, end-diastolic volume; and A, curve-fitting constant. ß was indexed to ventricular volumes to create a dimensionless index (ß, i). From

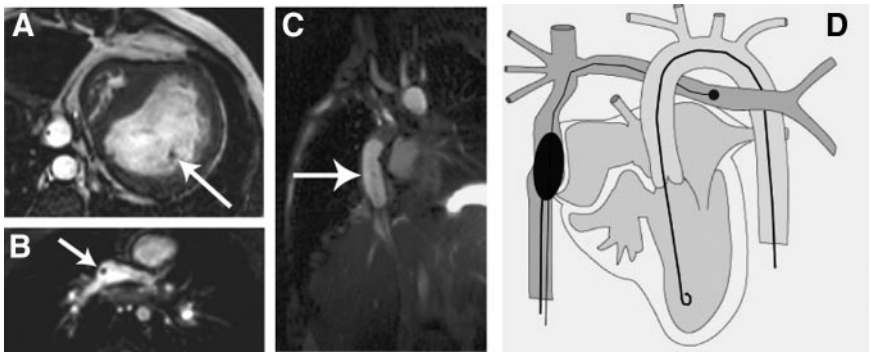


Figure 1. A, Cine MRI of catheters as used for ventricular pressure measurements. B, Cine MRI of pulmonary pressure measurement. The catheters are visible due to small local signal voids (arrows). C, An interactive real-time MRI of a balloon catheter in the inferior vena cava during inflation with isotonic saline solution. D, Illustration of the position of the catheters in the pulmonary artery, the single ventricle, and the inferior vena cava.

pressure measurements, we derived relaxation time constant τ as a parameter of early diastolic relaxation.

Pressure Recordings and Catheter Visualization During MRI

The fluid-filled catheters were connected to pressure transducers, and pressures were amplified, recorded, and analyzed with Ponemah software. An additional pressure transducer was positioned within the bore of the scanner to obtain a trigger signal for synchronizing measured pressures with cine and VEC MRI-derived ventricular and blood flow volumes.[21] The position of the catheters was visualized on cine MRI images (Figure 1). Appropriate inflation of the balloon catheter with saline solution was confirmed on interactive real-time MRI (Figure 1).[26]

Statistical Analysis

Measurements at rest and during dobutamine stress were analyzed with paired Student *t* test and Bonferroni-Holm correction for multiple comparisons of 20 parameters. Data are expressed as mean±SD. The correlation was determined among ventricular volumes, cardiac index, and aortopulmonary collateral flow. In addition, correlation was determined among measurements of collateral flow, PVR, and ergometry-derived parameters of functional capacity. The agreement between pulmonary and caval flow volumes was assessed using the Bland-Altman test.

Results

General Characteristics

We investigated 10 patients with a Fontan circulation (5 each with a right- and left-type single ventricle). Eight patients were white, and 2 were of Arabic ethnicity. All patients were compliant to the study. The duration of dobutamine exposure ranged from 15 to 25 minutes. There were no side effects to dobutamine. Due to the length of the study protocol, flow measurements in the inferior and superior vena cava were not performed in 2 patients at rest and in 4 patients during dobutamine stress.

All patients were New York Heart Association class I to II. Ergometry showed decreased functional capacity compared to published reference levels (Table 1).[27] Angiograms revealed no obstruction within the Fontan circulation and either no or only visibly small venovenous or aortopulmonary collaterals. Oxygen saturation at rest ranged from 91% to 97% and did not decrease significantly during dobutamine stress.

There were no statistically significant differences between parameters as measured for the left- versus right-type single ventricle. This finding also comprises the functional parameters as given in the next sections.

Blood Flow and PVR

The Nakata index of the pulmonary arteries was 150±45 mm^2/m^2 and, thus, substantially below published reference values of healthy controls.[1,18] Quantitative flow volumes measured in the inferior and superior vena cava were similar to those measured in the left and right pulmonary artery, and they all increased during dobutamine stress ($P<0.05$) (Table 2). In all patients,

Table 1. Patient Characteristics

	Systemic Ventricle	Age, y	Body Weight, kg	BSI, m^2	Myocardial MM, g	NYHA	Vo_2max, mL/min/kg	VE/Vco_2 Slope
Patient no. and leading condition								
1. Mitral atresia, TGA	RV	14	24	0.9	44	I	21.8	47
2. Tricuspid atresia	LV	16	27	1.0	61	I	22.4	36
3. Pulmonary atresia	LV	40	59	1.6	101	I	20.5	27
4. Unbalanced AV canal	LV	14	29	1.1	56	I	15.1	40
5. Tricuspid atresia, TGA	LV	39	69	1.9	103	II	16.9	28
6. DILV, TGA	RV	13	29	1.0	54	I	18	30
7. Univentricular heart	RV	26	86	1.9	97	I	21.6	29
8. Tricuspid atresia	LV	21	57	1.6	55	II	22.2	28
9. DORV, TGA	RV	18	39	1.3	76	II	20.1	34
10. DORV, heterotaxia	RV	18	56	1.5	88	I	14.9	33
Mean±SD		22±10	48±20	1.4±0.4	73±21		19.4±2.9	33.3±6.3

AV indicates atrioventricular; BSI, body surface index; DILV, double inlet left ventricle; DORV, double outlet right ventricle; LV, left ventricle; MM, muscle mass; NYHA, New York Heart Association; RV, right ventricle; TGA, transposition of the great arteries; VE/Vco_2 slope, minute ventilation to carbon dioxide production relationship; Vo_2max, peak oxygen uptake.

Table 2. Blood Flow Volumes

	SVC+IVC, L/min/m^2		LPA+RPA, L/min/m^2		Aorta, L/min/m^2		APC, L/min/m^2		APC Contribution to Qp, %	
	Rest	Dobu	Rest	Dobu	Rest	Dobu	Rest	Dobu	Rest	Dobu
Patient no.										
1	…	…	1.9	2.6	2.1	3.2	0.2	0.6	13.3	24.2
2	1.4	…	1.3	2.1	1.4	2.8	0.1	0.7	8.5	33.3
3	…	…	2.8	5.3	3.4	6.7	0.6	1.4	21.7	26.6
4	3.0	4.4	2.8	4.0	3.2	5.6	0.4	1.5	14.4	37.5
5	2.0	2.6	2.1	2.6	2.3	3.3	0.3	0.7	13.4	26.8
6	1.9	3.3	1.7	3.1	1.9	3.4	0.2	0.3	12.0	8.1
7	1.5	2.4	1.5	2.3	1.6	2.7	0.2	0.4	10.3	19.4
8	1.9	2.3	1.7	1.9	1.9	2.6	0.2	0.6	12.6	33.6
9	1.8	…	2.0	3.4	2.2	4.0	0.2	0.6	9.7	17.6
10	1.8	3.1	2.0	3.2	2.3	4.1	0.3	0.9	14.7	28.7
Mean±SD	1.9±0.4	3.0±0.7	2.0±0.5	3.1±1.0	2.2±0.6	3.8±1.3	0.3±0.1	0.8±0.4	13.1±3.5	25.6±8.3
P	0.0022*		0.0004*		0.0001*		0.0004*		0.0015*	

P values were adjusted by Bonferroni-Holm correction for multiple comparisons. Uncorrected *P* values were derived from paired Student *t* tests for differences between rest and dobutamine. APC indicates aortopulmonary collateral; Dobu, dobutamine; IVC, inferior vena cava; LPA, left pulmonary artery; Qp, pulmonary blood flow; RPA, right pulmonary artery; SVC, superior vena cava.

*Significant differences after Bonferroni-Holm correction.

there was only trivial retrograde flow in the pulmonary arteries both at rest and during stress (Figure 2). The Bland-Altman analysis showed good agreement between flow measurements obtained in the vena cava and the pulmonary arteries (Figure 3). During augmented pulmonary flow, mean pulmonary artery pressure and transpulmonary pressure gradient only changed slightly, and consequently, vascular resistance was decreased (Table 3). Blood flow volumes were significantly larger when measured in the aorta than when measured in the pulmonary arteries, and compared to rest, this difference increased during stress (Table 2). Thus, the calculated collateral blood flow increased significantly during dobutamine stress ($P<0.01$). There was no significant correlation between collateral blood flow and ventricular end-diastolic volumes (at rest, $r=-0.072$;

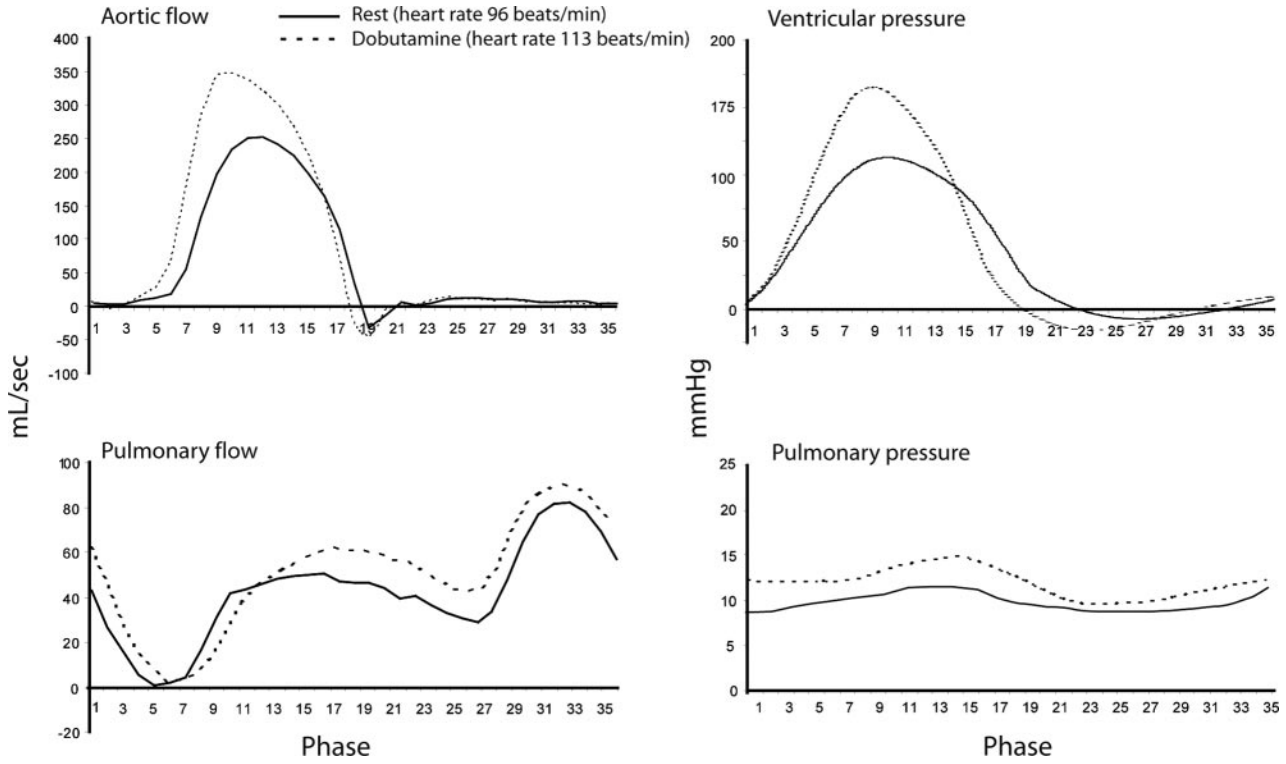


Figure 2. Representative flow volumes and pressures of a patient with Fontan measured at rest and during dobutamine stress. The pulmonary flow is shown as the sum of the flow measured in the left and right pulmonary artery. For illustration purposes, data measured at 2 different heart rates (rest and dobutamine stress) were scaled to 35 heart phases.

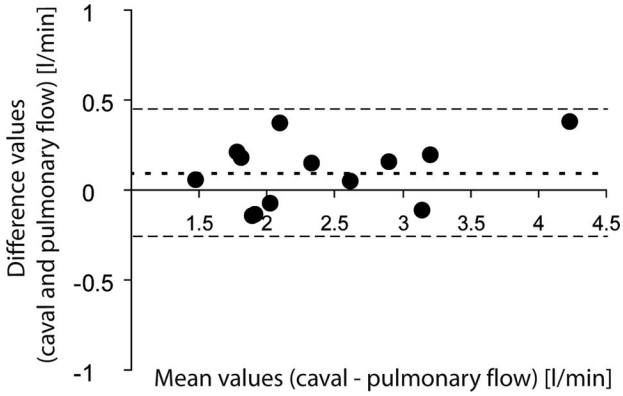


Figure 3. Bland-Altman plot of pulmonary and caval blood flow volumes (mL) for measurements obtained at rest and during dobutamine stress.

during dobutamine stress, $r=-0.062$) or between collateral blood flow and cardiac index (at rest, $r=-0.072$; during dobutamine stress, $r=-0.062$). However, there were statistically significant inverse correlations between collateral blood flow (aortopulmonary collateral) and both PVR ($r=-0.6$; $P=0.03$) and ergometry-derived peak oxygen uptake ($r=-0.7$; $P=0.01$). Blood flow patterns in the pulmonary arteries were similar at rest and during dobutamine stress (Figure 2). In concordance with other reports, the profiles showed some degree of interindividual variability.[28,29]

Global Pump Function

During dobutamine stress, heart rate and cardiac output increased significantly, but stroke volume did not (Tables 4 and 5). Differences in ejection fraction just fell short of being significant when using Bonferroni-Holm adjustments for multiple comparisons ($P=0.009$) (Table 5).

Myocontractile and Diastolic Function

During dobutamine stress, there was an increase of Emax,i that was not significant when adjusting for multiple comparisons ($P=0.026$) (Table 4). Efficiency of ventricular-arterial coupling did not improve during dobutamine stress due to a concomitant increase in Ea.[25] By echocardiography, the E/A-wave ratio was above 0.7 in all patients. During MRI catheterization at rest, the relaxation time constant τ was at published reference levels for healthy controls and shortened significantly during dobutamine stress (Table 4). At the same time, there was a decrease of end-diastolic volumes (Figure 4 and Table 5). End-diastolic pressure increased slightly but not significantly. There were no significant changes of the stiffness constant ß, but the pressure-volume loops shifted toward the left in the pressure-volume diagram in all patients (Figure 4 and Table 4).

Discussion

A progressive decrease of exercise capacity is commonly observed in patients with Fontan circulation, and at late follow-up, heart failure contributes importantly to morbidity.[5] The pathophysiologic causes for heart failure seem to be multifactorial. In this study, MRI catheterization was used at rest and during dobutamine stress to obtain information about PVR, global ventricular pump function, myocardial contractility, and diastolic function. The major findings are that left-to-right shunt through aortopulmonary collaterals increased during dobutamine stress, but PVR decreased. In addition, during stress, the single ventricle had signs for abnormal diastolic compliance.

Blood Flow and PVR

Several authors reported a diminished growth of the pulmonary arteries despite somatic growth and assumed that this might have an impact on PVR.[1,2] However, only sparse data are available

Table 3. Ventricular Pressures and Pulmonary Pressures and Resistance

	Ventricular Pressures				Pulmonary Pressures and Resistance					
	EDP, mm Hg		ESP, mm Hg		Mean PAP, mm Hg		TPG, mm Hg		PVR, Wood Units/BSI	
	Rest	Dobu	Rest	Dobu	Rest	Dobu	Rest	Dobu	Rest	Dobu
Patient no.										
1	2.2	4.1	90.0	110.0	9.7	11.7	7.5	7.6	3.6	2.4
2	6.0	8.2	98.0	104.0	10.6	12.8	4.6	4.6	3.2	1.7
3	7.7	7.9	82.1	101.0	11.3	12.9	3.6	4.2	1.0	0.6
4	4.6	5.6	80.0	97.0	9.0	13.3	4.4	7.7	1.4	1.4
5	4.0	6.2	105.0	103.0	10.1	11.7	6.1	5.5	2.6	1.7
6	5.8	5.9	91.6	85.0	10.0	13.2	4.2	6.4	2.2	1.9
7	5.5	8.3	78.0	97.0	12.7	12.9	7.2	4.3	4.5	1.6
8	5.6	7.1	96.0	111.0	12.2	13.3	6.6	5.8	3.5	2.3
9	6.3	7.0	91.0	117.0	12.5	12.0	6.2	5.0	2.9	1.2
10	8.3	9.0	93.1	119.0	12.4	12.9	4.1	3.5	1.8	0.8
Mean±SD	5.6±1.7	6.9±1.4	90.5±8.0	104.4±9.7	11.1±1.3	12.7±0.6	5.5±1.3	5.5±1.4	2.7±1.0	1.6±0.5
P	0.084		0.0034*		0.0057*		0.986		0.0022*	

P values were adjusted by Bonferroni-Holm correction for multiple comparisons. Uncorrected *P* values were derived from paired Student *t* tests for differences between rest and dobutamine. BSI indicates body surface index; EDP, end-diastolic pressure; ESP, end-systolic pressure; PAP, pulmonary arterial pressure; TPG, transpulmonary gradient.

*Significant differences after Bonferroni-Holm correction.

Table 4. Pulmonary Function

	Heart Rate, bpm		Contractility Emax, mm Hg/mL/100 g		Coupling Emax/Ea		Diastolic Compliance ß, 1/mL/100 mL EDV		Early Diastolic Relaxation τ	
	Rest	Dobu	Rest	Dobu	Rest	Dobu	Rest	Dobu	Rest	Dobu
Patient no.										
1	83	149	1.5	2.6	0.4	0.5	0.7	0.5	48.2	28.4
2	107	143	1.9	3.2	0.3	0.3	1.2	0.9	27.3	25.0
3	76	134	3.1	4.6	0.6	0.5	1.8	2.1	31.5	25.3
4	89	136	2.3	4.2	0.7	0.4	1.4	1.3	33.8	22.6
5	73	87	3.4	3.5	0.7	0.7	0.9	1.4	32.4	27.5
6	52	87	3.6	4.8	0.8	0.5	1.1	1.0	40.1	36.0
7	98	142	4.0	4.2	0.6	0.6	1.6	1.7	23.5	18.0
8	91	113	4.1	5.5	0.8	0.9	1.2	1.3	49.2	23.0
9	81	129	3.5	3.9	0.2	0.2	1.8	1.9	57.6	43.0
10	88	141	4.2	6.8	0.9	0.8	1.7	1.8	23.1	15.0
Mean±SD	84±14	126.1±21.6	3.2±0.9	4.3±1.1	0.6±0.2	0.5±0.2	1.3±0.4	1.4±0.5	36.7±11.1	26.4±7.8
P	0.00002*		0.0264		0.2171		0.5212		0.0023*	

P values were adjusted by Bonferroni-Holm correction for multiple comparisons. Uncorrected *P* values were derived from paired Student *t* tests for differences between rest and dobutamine. ß indicates stiffness constant indexed to 100 mL of end-diastolic volume; Dobu, dobutamine.

*Significant differences after Bonferroni-Holm correction.

about the functional status of the pulmonary arteries at rest and during stress partly because vascular properties like PVR cannot be evaluated using conventional techniques, such as thermodilution or the Fick method, in patients with a Fontan circulation. In the present setting, MRI catheterization is a unique tool that combines invasive pressures and VEC MRI-derived flow data that also account for collateral aortopulmonary blood flow.[3,16,28] For reliable assessment of the PVR in the Fontan circulation, which is very susceptible to changes in volume load, pulmonary pressures and blood flow were measured in our study simultaneously but not sequentially.

The MRI method used in this study for measuring collateral blood flow was adopted from the work published by Grosse-Wortmann et al[3] and Whitehead et al.[14] In these studies, collateral flow was estimated by 2 approaches. Simplified, collateral flow was defined as the differences between flow volumes measured in the ascending aorta and the 4 pulmonary veins or, alternatively, as the difference between flow volumes in the ascending aorta and the systemic venous return. The latter was determined in the descending aorta[3] or the inferior and superior vena cava.[14] In both studies, the different methods had overall good agreement. In the present study, we modified the approach of Whitehead and colleagues slightly by directly measuring pulmonary arterial flow instead of caval flow. Comparing pulmonary with caval flow volumes showed good agreement (Figure 3).

Table 5. Ventricular Volumes

	EDV, mL/m^2		ESV, mL/m^2		SV, mL/m^2		EF, %		Cardiac Index, L/min/m^2	
	Rest	Dobu	Rest	Dobu	Rest	Dobu	Rest	Dobu	Rest	Dobu
Patient no.										
1	86.4	77.8	49.5	48.7	36.9	29.1	42.7	37.4	2.9	2.9
2	83.9	77.2	48.2	41.5	35.7	35.7	42.5	46.3	3.8	5.1
3	72.3	66.1	49.4	43.3	22.9	22.9	31.6	34.6	1.7	3.1
4	45.3	36.3	22.3	13.4	23.0	22.9	50.7	63.1	2.0	3.1
5	77.9	69.3	53.3	43.7	24.5	25.6	31.5	36.9	1.8	2.2
6	46.3	37.6	20.1	12.9	26.3	24.6	56.7	65.6	1.4	2.1
7	38.0	38.2	19.5	9.2	18.5	22.6	48.6	71.0	1.8	3.2
8	71.5	61.7	43.4	31.2	28.1	30.5	39.3	49.5	2.6	3.5
9	89.9	71.5	54.1	40.0	28.0	31.5	39.8	44.1	2.3	4.1
10	48.6	43.8	25.7	18.0	22.9	25.9	47.1	59.0	2.0	3.6
Mean±SD	66±18.5	57.9±16.2	38.6±14	30.2±14.5	26.7±5.5	27.1±4.2	43.1±7.6	50.7±12.4	2.2±0.7	3.3±0.8
P	0.0004*		0.0001*		0.6877		0.0092		0.0002*	

P values were adjusted by Bonferroni-Holm correction for multiple comparisons. Uncorrected *P* values were derived from paired Student *t* tests for differences between rest and dobutamine. Dobu indicates dobutamine; EDV, end-diastolic volume; EF, ejection fraction; ESV, end-systolic volume; SV, stroke volume.

*Significant differences after Bonferroni-Holm correction.

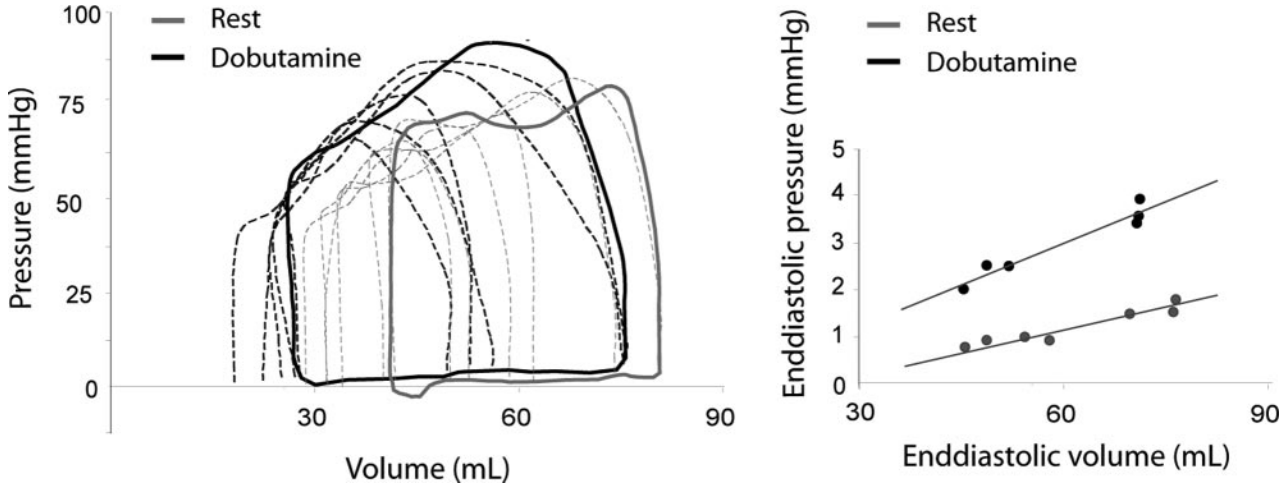


Figure 4. Pressure-volume loops (left) and end-diastolic pressure-volume points with regression line (right) in a representative patient at rest and during dobutamine stress. Note the left shift of the loops and end-diastolic pressure-volume points during dobutamine stress.

In concordance with the studies by Whitehead et al[14] and Grosse-Wortmann et al,[3] we noted an important contribution of aortopulmonary collateral flow to total effective pulmonary blood flow. The latter is considered being the amount of blood that passes through the pulmonary arteriole level, thus the sum of antegrade pulmonary and collateral flow. Recirculating blood through collaterals contributed about 13% to the total effective pulmonary flow at rest, and this percentage increased substantially to >25% during dobutamine stress. Interestingly, there was statistically significant inverse correlation between the amount of collateral flow and peak oxygen uptake during ergometry. Collateral flow also correlated inversely with PVR, but the relatively small sample size in the present study limits the clinical interpretation of this finding. This issue must be addressed systematically in future research.

In the studied patients, PVR at rest was slightly above published control values, which were determined in a previous study using the same MRI catheterization method.[16] During dobutamine stress, mean pulmonary pressures increased, but transpulmonary gradient did not, and thus, PVR decreased in the presence of augmented pulmonary throughput. PVR generally is considered a valid indicator of structural changes at the small resistive pulmonary arteriole level. Hence, we concluded that in the studied patients, there is at least a partially functioning pulmonary vasculature regulation to variation in pulmonary perfusion. However, one must keep in mind when interpreting these data that dobutamine was shown to increase pulmonary blood flow, but it does not affect the pulmonary vascular tone.[30,31] In addition, Hjortdal and colleagues[32] demonstrated that in patients with a Fontan circulation with total cavopulmonary connection, blood flow through the lung is driven by cardiac, respiratory, and peripheral muscular mechanisms. The relative contributions of these pumping mechanisms are influenced by many factors that include cardiac performance, physical exercise, body position, and respiratory patterns. These different mechanisms cannot be realistically simulated in the MRI environment.

Global Ventricular, Myocontractile, and Diastolic Function

Similar to earlier studies, we noted no substantial stroke volume augmentation under dobutamine stress in the patients with a Fontan circulation.[10,33] Myocardial contractility increased only slightly and not significantly during inotropic stimulation, and thus, the increase in cardiac output was mainly regulated by heart rate. This finding, again, is in line with earlier studies.[33] In conjunction with slightly improved parameters of contractility, end-systolic volumes decreased adequately, suggesting that systolic dysfunction is not the predominant cause of the abnormal response to stress. At the same time, end-diastolic volumes decreased, again similar to previous observations, and Emax/Ea, a parameter for the efficiency of ventriculoarterial coupling, remained at a low level.[10,23] These findings suggest an abnormal diastolic function of the single ventricle during dobutamine stress. In Fontan, diastolic filling is thought to be affected by a limited preload reserve because there is no subpulmonary ventricle actively pumping an appropriate amount of blood through the pulmonary system.[6,10,32] One also might speculate that in the presence of diminished preload recruitment, diastolic dysfunction is induced because the ventricle is not trained to be fully loaded. From elderly patients with atrial-septal defect, it is known that chronic underloading can lead to a small and stiff left ventricle.[34] In these patients, pharmacological priming before defect closure mostly improved diastolic function. On the contrary, recirculating blood through aortopulmonary collaterals contributes increasingly to ventricular filling under stress and, thus, would attenuate a deficient preload reserve. Other authors even discussed that collateral flow can impose a volume load on the single ventricle.[3,35,36] Overall, one should consider the important individual variations in the presence of collaterals and the resulting amount of collateral flow.[37]

In the present study, active early diastolic relaxation appeared to be normal, as indicated by a decrease of τ during dobutamine stress.[38] However, there was evidence of abnormal diastolic compliance. During dobutamine stress, the stiffness constant ß remained at similar levels as measured at rest, but the end-diastolic pressure-volume relation shifted toward the upper-left in

the pressure-volume diagram. Such a shift of the end-diastolic pressure-volume relation was not observed in recent studies that were performed in healthy pig left and right ventricles.[21,39] In addition, these studies showed that dobutamine has no major impact on the stiffness constant ß. It seems that chamber properties for filling do not improve during dobutamine stress; rather, filling must be accomplished in a system with increased stiffness. In the present study, we exposed the patients to moderate dobutamine stress that caused no substantial increase in end-diastolic pressures. In more severe stress situations, diastolic dysfunction might unmask even further, which could go along with more increased end-diastolic pressures. An abnormal ventricular stiffness could be explained by several reasons. The geometry of the single ventricle has a direct impact on its mechanical pump function and, thus, on its systolic and diastolic properties.[40] In addition, prolonged cyanosis and volume load during infancy are thought to induce myocardial fibrosis.[6,41] Finally, histopathologic studies in tricuspid atresia showed abnormal formation and arrangement of the fibrous matrix and the aggregation of myocyte chains.[11] Thus, the potential causes for the development of heart and pulmonary vascular dysfunction are multifactorial, as are the resulting forms of dysfunction, implying that it is essential to investigate these patients with methods that give a differential insight into the predominant form of failure that, in turn, will allow optimizing treatment concepts.

In summary, the findings of this study indicate that during dobutamine stress, blood flow through aortopulmonary collaterals contributes progressively to pulmonary perfusion. The pulmonary arteries had an abnormal growth index, but vascular regulation appeared to be unsuspicious. PVR decreased during stress and, thus, did not contribute to impaired diastolic filling in the patients studied. In contrast, we noted in the presence of normal early relaxation alteration of ventricular compliance during stress.

Limitations

There is no established animal model for studying the Fontan circulation. Additionally, extrapolating findings from healthy human controls to a univentricular physiology must be made with great care. Therefore, this patient study has a descriptive nature, and findings must be related to parameters that were obtained in other research. Care must be taken when comparing changes of ventricular and pulmonary vascular function induced by physical exercise with dobutamine stress.[30,32,42,43] Therefore, it would be inappropriate to directly translate our findings to exercise conditions. In addition, heart rate effects must be considered when interpreting our data of diastolic compliance. Collateral flow through intrapulmonary shunts and venovenous collaterals was not determined in our study. However, arterial oxygen saturation did not decrease significantly during stress, and thus, one can assume that flow through these types of shunts was relatively constant during the study protocol. Measuring blood flow in all 4 pulmonary veins would have allowed quantifying the amount of venovenous and intrapulmonary flow but at the expense of a substantially lengthened protocol that was judged too demanding. In general, VEC phase-contrast MRI flow sequences must be used with great care when measuring quantitative flow.

Finally, there is a broad range of varieties in Fontan regarding anatomic conditions and the onset and time course of cardiovascular dysfunction. Therefore, one cannot extrapolate our data of preselected patients to other forms of Fontan. However, the method used in this study allows differentiation among pulmonary, systolic, and diastolic dysfunction, which will potentially improve the planning of individual treatment strategies.

Sources of Funding

This work was supported by the German Science Association (DFG) and the German Federal Ministry of Education and Research (grant 01EV0704). Responsibility for the contents rests with the authors.

Disclosures

None.

References

1. Ovroutski S, Ewert P, Alexi-Meskishvili V, Holscher K, Miera O, Peters B, Hetzer R, Berger F. Absence of pulmonary artery growth after Fontan operation and its possible impact on late outcome. *Ann Thorac Surg*. 2009;87:826–831.
2. Tatum GH, Sigfusson G, Ettedgui JA, Myers JL, Cyran SE, Weber HS, Webber SA. Pulmonary artery growth fails to match the increase in body surface area after the Fontan operation. *Heart*. 2006;92:511–514.
3. Grosse-Wortmann L, Al-Otay A, Yoo SJ. Aortopulmonary collaterals after bidirectional cavopulmonary connection or Fontan completion: quantification with MRI. *Circ Cardiovasc Imaging*. 2009;2:219–225.
4. Spicer RL, Uzark KC, Moore JW, Mainwaring RD, Lamberti JJ. Aortopulmonary collateral vessels and prolonged pleural effusions after modified Fontan procedures. *Am Heart J*. 1996;131:1164–1168.
5. Khairy P, Fernandes SM, Mayer JE Jr, Triedman JK, Walsh EP, Lock JE, Landzberg MJ. Long-term survival, modes of death, and predictors of mortality in patients with Fontan surgery. *Circulation*. 2008;117:85–92.
6. Gewillig M. The Fontan circulation. *Heart*. 2005;91:839–846.
7. Cheung YF, Penny DJ, Redington AN. Serial assessment of left ventricular diastolic function after Fontan procedure. *Heart*. 2000;83:420–424.
8. Mahle WT, Coon PD, Wernovsky G, Rychik J. Quantitative echocardiographic assessment of the performance of the functionally single right ventricle after the Fontan operation. *Cardiol Young*. 2001;11:399–406.
9. Piran S, Veldtman G, Siu S, Webb GD, Liu PP. Heart failure and ventricular dysfunction in patients with single or systemic right ventricles. *Circulation*. 2002;105:1189–1194.
10. Senzaki H, Masutani S, Ishido H, Taketazu M, Kobayashi T, Sasaki N, Asano H, Katogi T, Kyo S, Yokote Y. Cardiac rest and reserve function in patients with Fontan circulation. *J Am Coll Cardiol*. 2006;47:2528–2535.
11. Sanchez-Quintana D, Climent V, Ho SY, Anderson RH. Myoarchitecture and connective tissue in hearts with tricuspid atresia. *Heart*. 1999;81:182–191.
12. Beerbaum P, Barth P, Kropf S, Sarikouch S, Kelter-Kloepping A, Franke D, Gutberlet M, Kuehne T. Cardiac function by MRI in congenital heart disease: impact of consensus training on interinstitutional variance. *J Magn Reson Imaging*. 2009;30:956–966.
13. Korperich H, Gieseke J, Barth P, Hoogeveen R, Esdorn H, Peterschroder A, Meyer H, Beerbaum P. Flow volume and shunt quantification in pediatric congenital heart disease by real-time magnetic resonance velocity mapping: a validation study. *Circulation*. 2004;109:1987–1993.
14. Whitehead KK, Gillespie MJ, Harris MA, Fogel MA, Rome JJ. Noninvasive quantification of systemic-to-pulmonary collateral flow: a major source of inefficiency in patients with superior cavopulmonary connections. *Circ Cardiovasc Imaging*. 2009;2:405–411.
15. Kuehne T, Saeed M, Reddy G, Akbari H, Gleason K, Turner D, Teitel D, Moore P, Higgins CB. Sequential magnetic resonance monitoring of pulmonary flow with endovascular stents placed across the pulmonary valve in growing swine. *Circulation*. 2001;104:2363–2368.
16. Kuehne T, Yilmaz S, Schulze-Neick I, Wellnhofer E, Ewert P, Nagel E, Lange P. Magnetic resonance imaging guided catheterisation for assessment of pulmonary vascular resistance: in vivo validation and clinical application in patients with pulmonary hypertension. *Heart*. 2005;91:1064–1069.
17. Naeije R. Pulmonary vascular resistance. A meaningless variable? *Intensive Care Med*. 2003;29:526–529.
18. Nakata S, Imai Y, Takanashi Y, Kurosawa H, Tezuka K, Nakazawa M, Ando M, Takao A. A new method for the quantitative standardization of

cross-sectional areas of the pulmonary arteries in congenital heart diseases with decreased pulmonary blood flow. *J Thorac Cardiovasc Surg*. 1984;88:610–619.
19. Bodhey NK, Beerbaum P, Sarikouch S, Kropf S, Lange P, Berger F, Anderson RH, Kuehne T. Functional analysis of the components of the right ventricle in the setting of tetralogy of Fallot. *Circ Cardiovasc Imaging*. 2008;1:141–147.
20. Sarikouch S, Peters B, Gutberlet M, Leismann B, Kelter-Kloepping A, Koerperich H, Kuehne T, Beerbaum P. Sex-specific pediatric percentiles for ventricular size and mass as reference values for cardiac MRI: assessment by steady-state free-precession and phase-contrast MRI flow. *Circ Cardiovasc Imaging*. 2010;3:65–76.
21. Schmitt B, Steendijk P, Lunze K, Ovroutski S, Falkenberg J, Rahmanzadeh P, Maarouf N, Ewert P, Berger F, Kuehne T. Integrated assessment of diastolic and systolic ventricular function using diagnostic cardiac magnetic resonance catheterization: validation in pigs and application in a clinical pilot study. *JACC Cardiovasc Imaging*. 2009;2:1271–1281.
22. Brimioulle S, Wauthy P, Ewalenko P, Rondelet B, Vermeulen F, Kerbaul F, Naeije R. Single-beat estimation of right ventricular end-systolic pressure-volume relationship. *Am J Physiol Heart Circ Physiol*. 2003;284:H1625–H1630.
23. Kuehne T, Yilmaz S, Steendijk P, Moore P, Groenink M, Saaed M, Weber O, Higgins CB, Ewert P, Fleck E, Nagel E, Schulze-Neick I, Lange P. Magnetic resonance imaging analysis of right ventricular pressure-volume loops: in vivo validation and clinical application in patients with pulmonary hypertension. *Circulation*. 2004;110:2010–2016.
24. Steendijk P, Meliga E, Valgimigli M, Ten Cate F, Serruys PW. Acute effects of alcohol septal ablation on systolic and diastolic left ventricular function in patients with hypertrophic obstructive cardiomyopathy. *Heart*. 2008;94:1318–1322.
25. Burkhoff D, Sagawa K. Ventricular efficiency predicted by an analytical model. *Am J Physiol*. 1986;250:R1021–R1027.
26. Krueger JJ, Ewert P, Yilmaz S, Gelernter D, Peters B, Pietzner K, Bornstedt A, Schnackenburg B, Abdul-Khaliq H, Fleck E, Nagel E, Berger F, Kuehne T. Magnetic resonance imaging-guided balloon angioplasty of coarctation of the aorta: a pilot study. *Circulation*. 2006;113:1093–1100.
27. Dubowy KO, Baden W, Bernitzki S, Peters B. A practical and transferable new protocol for treadmill testing of children and adults. *Cardiol Young*. 2008;18:615–623.
28. Muthurangu V, Razavi RS. The value of magnetic resonance guided cardiac catheterisation. *Heart*. 2005;91:995–996.
29. Hager A, Fratz S, Schwaiger M, Lange R, Hess J, Stern H. Pulmonary blood flow patterns in patients with Fontan circulation. *Ann Thorac Surg*. 2008;85:186–191.
30. Pagnamenta A, Fesler P, Vandinivit A, Brimioulle S, Naeije R. Pulmonary vascular effects of dobutamine in experimental pulmonary hypertension. *Crit Care Med*. 2003;31:1140–1146.
31. Morita S, Kormos RL, Astbury JC, Shaub RD, Kawai A, Griffith BP. Standardized ejection fraction as a parameter of overall ventricular pump function. *Jpn Circ J*. 2000;64:510–515.
32. Hjortdal VE, Emmertsen K, Stenbog E, Frund T, Schmidt MR, Kromann O, Sorensen K, Pedersen EM. Effects of exercise and respiration on blood flow in total cavopulmonary connection: a real-time magnetic resonance flow study. *Circulation*. 2003;108:1227–1231.
33. Robbers-Visser D, Jan Ten Harkel D, Kapusta L, Strengers JL, Dalinghaus M, Meijboom FJ, Pattynama PM, Bogers AJ, Helbing WA. Usefulness of cardiac magnetic resonance imaging combined with low-dose dobutamine stress to detect an abnormal ventricular stress response in children and young adults after Fontan operation at young age. *Am J Cardiol*. 2008;101:1657–1662.
34. Schubert S, Peters B, Abdul-Khaliq H, Nagdyman N, Lange PE, Ewert P. Left ventricular conditioning in the elderly patient to prevent congestive heart failure after transcatheter closure of atrial septal defect. *Catheter Cardiovasc Interv*. 2005;64:333–337.
35. Bradley SM, McCall MM, Sistino JJ, Radtke WA. Aortopulmonary collateral flow in the Fontan patient: does it matter? *Ann Thorac Surg*. 2001;72:408–415.
36. McElhinney DB, Reddy VM, Tworetzky W, Petrossian E, Hanley FL, Moore P. Incidence and implications of systemic to pulmonary collaterals after bidirectional cavopulmonary anastomosis. *Ann Thorac Surg*. 2000;69:1222–1228.
37. Baile EM, Ling H, Heyworth JR, Hogg JC, Pare PD. Bronchopulmonary anastomotic and noncoronary collateral blood flow in humans during cardiopulmonary bypass. *Chest*. 1985;87:749–754.
38. Zile MR, Baicu CF, Gaasch WH. Diastolic heart failure-abnormalities in active relaxation and passive stiffness of the left ventricle. *N Engl J Med*. 2004;350:1953–1959.
39. Machii T, Yokota M, Nagata K, Ishihara H, Iwase M, Sobue T. Effect of dobutamine and OPC-18790 on diastolic chamber stiffness in patients with idiopathic dilated cardiomyopathy. *J Cardiovasc Pharmacol*. 1997;29:265–272.
40. Akagi T, Benson LN, Gilday DL, Ash J, Green M, Williams WG, Freedom RM. Influence of ventricular morphology on diastolic filling performance in double-inlet ventricle after the Fontan procedure. *J Am Coll Cardiol*. 1993;22:1948–1952.
41. van den Berg J, Wielopolski PA, Meijboom FJ, Witsenburg M, Bogers AJ, Pattynama PM, Helbing WA. Diastolic function in repaired tetralogy of Fallot at rest and during stress: assessment with MR imaging. *Radiology*. 2007;243:212–219.
42. Oosterhof T, Tulevski II, Roest AA, Steendijk P, Vliegen HW, van der Wall EE, de Roos A, Tijssen JG, Mulder BJ. Disparity between dobutamine stress and physical exercise magnetic resonance imaging in patients with an intra-atrial correction for transposition of the great arteries. *J Cardiovasc Magn Reson*. 2005;7:383–389.
43. Kafi SA, Melot C, Vachiery JL, Brimioulle S, Naeije R. Partitioning of pulmonary vascular resistance in primary pulmonary hypertension. *J Am Coll Cardiol*. 1998;31:1372–1376.

CLINICAL PERSPECTIVE

The pathophysiologic causes for failure of the Fontan circulation are multifactorial. Therefore, diagnostic tools are warranted that permit a differential analysis of ventricular and pulmonary vascular function. In this study, an MRI catheterization technique that enables simultaneous pressure and volume measurement in the single ventricle was used. From these measurements, parameters of global pump, myocontractile, and diastolic function can be derived. In addition, MRI catheterization allows determination of aortopulmonary collateral flow in conjunction with pulmonary vascular resistance. We found that pharmacological stress by dobutamine improved contractility, although without substantial augmentation of stroke volumes. At the same time, the single ventricle showed signs of abnormal diastolic performance. In the absence of a subpulmonary ventricle, these findings should be seen in the light of pulmonary vascular function. In the studied patients, blood flow through aortopulmonary collaterals contributed substantially to the total pulmonary blood flow. In addition, its proportion increased during stress. However, augmented total pulmonary blood flow was not associated with increased pulmonary vascular resistance, implying that resistance did not contribute to a limited preload reserve and, thus, impaired diastolic filling of the systemic ventricle. The method described in this study provides detailed and differential information of the cardiovascular function in Fontan, which will potentially improve the planning of individual treatment strategies. The findings of this descriptive study of preselected patients require further study in larger groups of patients with different types of Fontan circulation.

Psychological Trauma: Theory, Research, Practice, and Policy
2010, Vol. 2, No. 1, 54–62

1942-9681/10/$12.00 DOI: 10.1037/a0019020

Challenges to Engaging Black Male Victims of Community Violence in Healthcare Research: Lessons Learned From Two Studies

Sonia Schwartz
Boston Medical Center and Boston University School of Medicine

Joel Hoyte
Tufts University

Thea James
Boston Medical Center

Lauren Conoscenti
National Center for PTSD, VA Boston Healthcare System

Renee Johnson
Boston University School of Public Health

Jane Liebschutz
Boston Medical Center and Boston University School of Public Health

A dearth of literature exists on barriers to conducting research with Black male victims of community violence, despite the need for evidence-based postinjury interventions. This study used qualitative data from a cross-sectional interview study ($n = 16$) and a pilot intervention study ($n = 11$) conducted in Boston, MA to identify challenges and facilitators to conducting research with Black male victims of community violence, particularly with regard to recruitment and maintenance of a study sample. Qualitative methods, including Grounded Theory and ethnography, were used to analyze the data. Challenges included a fear of police involvement, an impression of "snitching" when disclosing personal information, mistrust of research motives, suspicion of the informed consent process, the emotional impact of the trauma itself, and logistical issues. Facilitators to research included monetary incentives and motivation to help oneself and others. Participant recommendations on recruitment methods relating to approach and timing are provided. Findings from this study may assist in the planning of research studies for Black male victims of community violence.

Keywords: African American, qualitative research, community violence, research participation

Morbidity and mortality from stabbing and shooting violence disproportionately affects young Black males. In 2006, the firearm homicide rate for Black males ages 18–25 was 100.4 per 100,000 persons, more than five times higher than the rate for any other group (National Center for Injury Prevention and Control, 2005). Despite potential for social, behavioral, and other hospital-based interventions to improve the treatment received by Black male victims of community violence, relatively few studies have been conducted in these settings (Becker, Hall, Ursic, Jain, & Calhoun, 2004; Cooper, Eslinger, Nash, al-Zawahri, & Stolley, 2000; Cooper, Eslinger, & Stolley, 2006; Zun, Downey, & Rosen, 2006). Moreover, little research is available to describe the specific barriers and facilitators to research participation in this population.

Two bodies of literature can help to inform this research gap. The first addresses conducting research with African Americans. This topic has been examined in a variety of samples, including cancer patients, general medical patients, and community members. Studies cite a lack of trust in medical research, particularly stemming from the Tuskegee Study, as a central barrier to conducting research with African Americans (Corbie-Smith, Thomas, Williams, & Moody-Ayers, 1999; McCallum, Arekere, Green, Katz, & Rivers, 2006; Rajakumar, Thomas, & Musa, 2009; Shavers, Lynch, & Burmeister, 2002). Other barriers include the belief that minorities have more risks in medical research, fear of risks to participating, inconvenience, perceived physician dishonesty, lack of perceived need for research, belief that African Americans would not benefit from any advancements a study may bring, fear of worsening health that may result in participation, and confusion of the purpose and meaning of informed consent (Corbie-Smith, Thomas, & St. George, 2002; Corbie-Smith et al., 91999; Dunlop,

Sonia Schwartz, Section of General Internal Medicine, Boston Medical Center, Clinical Addiction Research and Education (CARE) Unit and Department of Medicine, Boston University School of Medicine; Joel Hoyte, Department of Community Health, Tufts University; Thea James, Department of Emergency Medicine, Boston Medical Center; Lauren Conoscenti, National Center for PTSD, VA Boston Healthcare System; Renee M. Johnson, Boston University School of Public Health and Department of Community Health Sciences; Jane Liebschutz, Department of Social and Behavioral Sciences, Boston University School of Public Health and Section of General Internal Medicine, Boston Medical Center, Clinical Addiction Research and Education (CARE) Unit and Department of Medicine.

Correspondence concerning this article should be addressed to Jane Liebschutz, Department of Social and Behavioral Sciences, Boston University School of Public Health and Section of General Internal Medicine, Boston Medical Center, Clinical Addiction Research and Education (CARE) Unit and Department of Medicine, 801 Massachusetts Avenue, 2nd Floor, Boston, MA 02118. E-mail: jane.liebschutz@bmc.org

Graham, Leroy, Glanz, & Dunlop, 2007; Shavers, Lynch, & Burmeister, 2002).

By contrast, there is much less published work on the second body of literature, barriers to research participation among community violence survivors. A study of a pharmacological intervention in acute trauma patients, including victims of violence as well as survivors of motor vehicle crashes and other accidents, showed difficulties enrolling trauma patients from the hospital; of the 569 accessible and potentially eligible patients, only 48 (8%) enrolled (Stein, Kerridge, Dimsdale, & Hoyt, 2007). The reasons for this difficulty in enrollment have not been explored, although informal discussions with patients by Stein and colleagues revealed several possible explanations, such as patients not wanting to participate in anything that may delay their discharge from the hospital and denial of possible mental health outcomes from a traumatic event (Stein et al., 2007). In a randomized controlled trial of victims of violent crime, only 243 (11.2%) of the 2161 patients responded to a letter sent by the study team (Rose, Brewin, Andrews, & Kirk, 1999). Investigators in that study felt that patients' unwillingness to share their experience of victimization with strangers may have played a large role in this low response rate.

Although these studies suggest issues that may be important when designing studies for Black male victims of violence, research is needed to examine whether there are unique barriers not previously identified, and how much previously identified issues impact research participation among victims of violence. To address these research gaps, we culled data from two studies to identify the challenges and facilitators to conducting interview and intervention studies with Black male victims of community violence. One was a cross-sectional qualitative study on interactions with health care after injury. The other study was a pilot intervention to link Black male victims of community violence to culturally appropriate mental health and primary care.

Methods

Participants

Participants from the cross-sectional study were recruited via flyers posted in community settings around Boston, MA through word of mouth from other participants, and via online posting (boston.craigslist.org). Interested people were screened for eligibility criteria, which included male gender, age 18–40 years, and history of a gunshot or stab wound injury for which they sought medical care. Although Black race was not an eligibility criterion, all participants who completed the study self-identified as Black. Thus, participants for the cross-sectional study included 16 Black males with a median age of 31 years (range 25–38). Three were victims of gunshot, five of stab wounds, and eight were victims of both injuries. The median length of time since the most recent injury was 5.5 years (range 4 months to 20.1 year). Eligible participants provided written informed consent. At interview completion, they received $25 compensation and a referral list for local mental health and substance use treatment centers.

The second study was a pilot intervention that linked participants to primary care and provided a five-session problem-solving counseling program. Participants in the intervention study were recruited through a violence intervention advocacy program of an urban academic medical center with the largest Level 1 trauma center in New England. Eligibility criteria included gunshot or stab wound that was treated via the Emergency Department in the prior 2 years, age ≥18 years, self-identified Black race, male sex, English language fluency, plans to stay in the area for the next 6 months, two forms of contact, and use of an illicit substance or hazardous amounts of alcohol in the prior month. Those who suffered a traumatic brain injury were excluded because of the need for participants to be able to engage in the counseling program. Participants for the intervention included 11 self-identified Black males with median age of 26 years (range 18–42). Two had been treated for a stab wound and nine for a gunshot injury that occurred a median of 23 days (range 1 day to 23 months) before study entry. One participant enrolled in both studies. Enrolled participants provided written informed consent and received $25 compensation for each of three research interviews completed.

Procedures

In the cross-sectional study, trained interviewers conducted 14 semistructured interviews from January through December of 2008 (2 paired and 12 individual interviews). Seven interviews were conducted by 2 Black, male, community-based mental health professionals, 3 by a White female research assistant, 3 by a Black male research assistant, and 1 by the White female principal investigator. To facilitate rapport, all interviews were scheduled with Black, male interviewers. However, two interviews occurred when participants showed up hours past the scheduled interview time, when the designated interviewer was no longer available. Two interviews occurred when the participants could not provide contact information to allow confirmation with Black interviewer schedules and thus were interviewed right at the time of inquiry by available trained staff. Interviews were digitally recorded and lasted 45−120 min. Audio recordings were professionally transcribed and identifying information was changed to preserve participant confidentiality.

The interview guide focused largely on experiences with health care during and after injuries. Once the investigators encountered recruitment difficulties, questions were added about views on research participation, particularly recruitment, starting with the third participant. The goals were to inform improvements to the recruitment protocol and to anticipate research conduct issues that might arise in the planned intervention study. Participants in later interviews were asked for clarification on concepts discussed in earlier interviews or about issues that came up during initiation of the intervention study.

In the pilot intervention study, data were drawn from detailed ethnographic field notes (Warren & Karner, 2005) taken by the research assistant on all contacts with participants over the course of their participation in this 6-month study. Notes explicitly focused on the process of research participation that could help assess feasibility of the program for future application. This included interactions relating to appointment scheduling, communication, family, transportation, and disability, as well as commentary on research from the participants.

Each study obtained approval from the Institutional Review Board at Boston University Medical Center and received Certificates of Confidentiality from the National Institutes of Health.

Analysis

Grounded Theory methods (Glaser & Strauss, 1967; Strauss & Corbin, 1990) were used to analyze the data, wherein themes found in earlier interviews were asked about in subsequent interviews for clarification if they were not spontaneously mentioned. The racially and professionally diverse research team represented a range of viewpoints for data interpretation. It included 3 White members and 3 Black members, and consisted of 2 physicians, 2 college-educated research assistants, 1 clinical psychologist, and 1 professor of public health. Team members listened to and read over the interview transcripts from the cross-sectional study and the field notes from the pilot intervention multiple times to create a coding scheme. At least two members coded each cross-sectional interview using NVIVO v. 7 software (QSR International Pty Ltd), resolving coding discrepancies through discussion with another team member. The pilot intervention field notes were coded by hand by at least two team members, resolving discrepancies with another team member. A full coding scheme was created from the cross-sectional interviews before analyzing the complete data from the pilot intervention. However, the field notes offered additional themes and ideas, and were integrated with the cross-sectional interviews. In some cases, the qualitative interviews explained observations made of the intervention. Questions on interpretation of data were clarified weekly with two violence intervention advocates, including one with a history of a gunshot wound. Throughout both studies, the team practiced reflexivity (Maruta, Swanson, & Finlayson, 1979) by continuously attending to the feelings and biases that emerged for them throughout the research process.

Results

Analysis of data from the cross-sectional and pilot intervention studies revealed challenges and facilitators to conducting research with Black male victims of community violence. Themes related to challenges, facilitators, and recommendations are described below along with the social context in which they occurred. Challenges to research centered on mistrust of the research process in several contexts, including fear of police involvement, an impression of "snitching" when revealing personal information, mistrust of research motives, and suspicion of the informed consent process. Other challenges included the emotional impact of the trauma itself as well as logistical issues. Facilitators to research included monetary incentives and motivation to help oneself and others. Participants' recommendations for recruitment are also presented. Unless otherwise noted, data came from the cross-sectional interview study. See Table 1 for a summary of these findings.

Challenges

Fear of police involvement. When asked about barriers to research participation, 7 (44%) of the 16 cross-sectional study participants voiced concern about police involvement in the research process. In particular, they feared that police could obtain information from the study to use against the participant or make a case in the injury investigation. One participant commented, "Like . . . maybe the undercover police, they're trying to find out the whole situation, what's going on with the story. Or, they might feel that you may give this story to the police." Another stated, "People think, 'Oh, they're trying to get some information from me. To get me caught up with the police or something like that."

One participant attributed this concern of being approached for research while hospitalized to the timing of the police investigation relating to his injury, ". . . 'cause most likely, the police usually come right after an injury. Then you guys come, so therefore, it's like, 'Well, the police just came. They might be connected to the police, too.'" This suspicion can lead some people to refuse speaking with anyone, as indicated by a participant who said: "I'm not talking to

Table 1
Recommendations for Conducting Research With Black Male Victims of Community Violence

Challenge	Recommendation
Fear of police involvement	Include on recruitment material: "Information will not be shared with police"; specify that you are not affiliated with police during initial approach and provide detailed information about the study to participant
Impression of snitching	Make privacy rules clear by reviewing rules again after consent, particularly relating to tape recorder or use of information
Mistrust of research motives	Provide information on purpose of research, funding and benefit to researchers Initial point of contact should be someone of similar background to potential participant-age, race, sex
Confusion of informed consent language (Principal) Investigator	Substitute with "research team member," "research director"
Privacy laws	Talk to IRB about ways to clarify language in consent form or through a consent script to make it easier to understand; explain specifically how data and forms are kept private—"This consent form will not go in the file with the other information you give us." or "We do not use your name on any of the forms other than the consent form"
Not open to discussing injury	Use mellow approach methods to first gauge the potential participants' willingness to talk
Logistical problems	Flexible scheduling, expect that many appointments will be missed, will start late, or will have to be rescheduled Provide taxi vouchers Use separate study telephone not attached to hospital line

Note. $n = 16$ for the cross-sectional study and $n = 11$ for the pilot intervention. The confusion of informed consent language, privacy laws, and logistical problems are informed by the cross-sectional and pilot intervention study. All other points are informed by the findings in the cross-sectional study.

anybody . . . "cause I don't know who they are, where they're from and what they're about."

Snitching. Five (31%) participants in the cross-sectional study spontaneously mentioned the perception that talking to researchers is a form of "snitching" or "ratting" on somebody. This included the fear that information would be shared outside of the research setting, and that participants were being "set up"; responses indicated misunderstanding the rules of privacy. One participant said:

> You don't want to talk about it. You get injured, you're like, "I'm not snitching. I'm not saying nothing to NOBODY. So, just talking to any study, or whatever, whatever, even though all the rules of the . . . of the study is . . . explain it's confidential and all that, but at the same time, some people just don't have their mind together and they think it's a form of snitching.

The use of recorders in the interviews played a role in the concern for information getting released, "Then you got a tape recorder right there, I'm like, 'Yo, what's that?' I don't want to swear but what the "f" are you going to do with that?"

Mistrust of research motives. Participants also noted that the motivation behind research was suspicious as exemplified by the language used to describe research, such as "there's always a catch" and "conspiracy." Two (13%) cross-sectional study participants identified research with lab animals, "You can't sit there and treat people from the inner city like a bunch of lab rats in a tank." Three (19%) participants discussed the belief that researchers have a "financial gain" in conducting studies:

> "The results of this research may be published in a medical book by White people for White people, to further benefit from your misery." That's what [the consent form] should state, like, [the research assistant] . . . is conducting this study, which is a LIE! So, why wouldn't she? She's . . . she's from [a hospital], right? With the rest of the White people that profit from all this misery? Very beneficial business.

This "profit" was assumed to benefit only the researchers, and not the participants or community, "Okay, it's a research study, they're gonna take my information and probably benefit, you know, from me, on a financial gain, and, you know, what am I gonna get out of it?"

Informed consent. Participants in both studies doubted the privacy protections provided by informed consent. It is important to note that all participants in both studies had undergone an informed consent process before the interview. One participant in the cross-sectional study said:

> 'Cause some people can say, "Yeah, this is a consent form, and we're not gonna give any information back," but then some people have different doubts, and they really think it's a situation like, "Well, I don't know, they might be trying to set me up in this situation."

Language of the consent form played a large role in this misunderstanding. During the pilot intervention, one participant signed the consent form. Later, he showed his copy to his mother. The next time the research assistant contacted the participant, he explained that his mother was concerned that his information would not be kept private and confidential AFTER she read the consent form, particularly the clauses that describe when researchers would be required to report private information. The research assistant explained the privacy laws again to both him and his mother, and the participant decided he would continue to participate. Additionally, the peer advocates working with the researchers noted that some study participants visibly tensed when hearing the words "principal investigator" read by the research assistant, as this term sounds similar to law enforcement terminology.

Not open to discussing injury. Another challenge to getting victims of community violence to participate in research relating to their injury was the emotional difficulty of talking about the injury. This concept was mentioned by 5 (31%) cross-sectional study participants. In the aftermath of the injury, victims of violence may understandably need time to sort out their feelings about the events. Feelings of fear, anger, disrespect, shock, and trauma were all mentioned as reasons why one may not want to participate in research soon after an injury, "After something happens to a person, depending on what happened and circumstances, it takes awhile for them to really get over the shocking point of it. You know what I mean? So it's like . . . 'give me time to BREATHE!'"

Participants mentioned the importance of taking into account the situations they encountered during the injury episode when approaching them about a study related to their injury, as this may affect their willingness to talk to unfamiliar researchers about what happened, "Like, Lord knows what might have happened when that person got shot. Like their best friend could have got killed and they just got injured, and they don't want to talk to nobody."

Because many victims of community violence are injured at a young age, one participant suggested that willingness to talk about the injury may also relate to the maturity level of the person at the time of the incident:

> Like, back then, I wouldn't have had time because I would rather be playing, than sitting up with somebody and talking for an hour. Like now, like, I'm grown, now. Like, I could sit here and conversate with you for an hour.

In addition to not wanting to discuss an injury right after it occurred, 4 (25%) of the cross-sectional study participants talked about the emotional pain that can come from bringing up the events of an injury that may have occurred several months or years ago, "You know, it hurts some people to think about it, you know? And their mind starts having flashbacks, you know? I'm dealing with them . . . memories that you're trying to suppress."

Logistics to research execution. During the intervention study, the field notes demonstrated a series of logistical challenges to working with this population, including scheduling appointments, telephone communication, and transportation. Appointments were scheduled to fit the needs of participants. Two (18%) of the 11 intervention participants had reversal of a day-night sleep cycle because of issues such as nightmares or an unstructured lifestyle, so they could only meet in the afternoons. Four (36%) intervention participants did not like to get home close to, or after, dark for fear of safety, so they requested morning appointments. Two (18%) intervention participants had childcare responsibilities and could only meet at certain times or days. Finally, on numerous occasions, participants did not show up at all or called at the time of the original appointment to change the time.

Telephone communication proved to be difficult throughout both studies. Participants rarely answered telephone calls when the caller identification showed a hospital telephone number. To address this issue, the study team purchased a cellular telephone for

the study, and noticed an immediate increase in the number of calls answered and returned. Of the 20 unanswered calls made using the hospital telephone for the intervention, just 4 (20%) were returned. By contrast, 4 out of the 9 (44%) unanswered calls from the study cellular telephone were returned. In addition, the hospital database of contact information was often incorrect. When asked about this in the cross-sectional study, one participant explained that providing inaccurate contact information reflected the "G-code," signifying a "gangster" code.

> That's . . . that's like sticking to what they call in the street, is like sticking to the "G Code." They . . . people just wanna get . . . get help that they need or whatever. But they don't want to . . . they don't want to participate in stuff, like the studies like this because they think this is like a form of snitching.

This participant explained that patients will be polite while getting treatment but may give a false telephone number to prevent any follow up contact.

Another issue stemming from telephone communication was the inability of some participants to maintain cellular telephone minutes. Three (27%) out of the 11 intervention participants lost their jobs after their injury and were unable to continue paying for their minutes. Four (36%) never had a steady income and it was unclear when cellular telephone minutes would be added back to the telephone. A second contact number was always taken at the screening for both studies, however, participants were not always available at the second contact, did not regularly stay at the location of the telephone, or it was the number of a person they might or might not see during the course of the study, like a parent or case manager.

A third logistical issue was transportation difficulty. Few participants had their own vehicles. All participants in the pilot intervention and 3 out of 16 (19%) in the cross-sectional study noted the fear of taking public transportation. At least one participant in each study was injured while on public transportation, and others were aware of the potential vulnerability while taking public buses or subways. One participant in the cross-sectional study describes his reaction to the police wanting him to take the bus down to the courthouse to testify as a witness:

> But, you know that, after . . . after something like this happen, there's no way you expect me just to get on the regular public transportation and come on down to the courthouse. That's unrealistic. That was like a death trap right there.

Physical disability can also affect transportation. Among the 11 intervention participants, 1 was paraplegic because of a spine injury, while 6 (55%) had leg injuries. This made traveling by public transportation or walking to the appointments challenging. To overcome both issues, taxi vouchers were provided in the pilot intervention for travel to and from appointments. Positive feedback was received from the pilot intervention participants regarding the vouchers, and none missed their medical appointments.

Factors Facilitating Research

Monetary incentives. When inquiring about facilitators to recruiting victims of violence for research, 7 (44%) of the 16 cross-sectional participants responded with some type of incentive, money being the most common. One participant stated, "I wouldn't do it if it was for free. I'd be like, 'I'm not gonna waste an hour of my time to like tell somebody about getting shot, and like get nothing out of the deal.'" Examples of nonmonetary incentives included counseling referrals ("So, you want to advertise, throw that out and say, 'Well, so, for our research, we'll also seek you help.'"), gift cards to toy stores ("I know that money's gonna go to my daughter"), movie passes, and food.

Participants mentioned not having a place to live, financial need, school loans, and the economy as some of the reasons for the importance of monetary incentives, "And it's sad to say that $25 might be life-altering at one time, another. They might need that money to go to work, or day labor, or whatever have you." One talked about the importance of compensation because it makes a person feel like they are participating in something meaningful: ". . . the money gives them the incentive to really feel like they're doing something."

Motivation to help oneself. Another reason why participants wanted to do research was they felt it would be beneficial to their own life. According to one participant, "You don't know what you might get out of it." Four (25%) cross-sectional participants believed that accessing resources through an intervention study, particularly counseling, would be a good incentive to participate in research.

> However, when you have a lot of people around you with a lot of feelings, you know what I mean, a lot of support . . . you know what I mean, it's gonna bring you to that point where you could get, you know, get . . . a little bit out of it. I'm not, I'm not saying all of it; a little bit out of it.

Conversely, not recognizing the personal benefits of research was also mentioned as a reason why one may not participate in a study. One participant said, "And I think most people'll go, 'Oh, f**k it. I ain't gonna go in there and tell them my God d**n life. Forget that. No, no, no. They, they ain't gonna do nothing, no way.'" Another explained, ". . . it's like, 'Well, I need my self-helped out. So, why would I want to help this person out if I can't get helped out?'"

Motivation to help others. Six (38%) cross-sectional participants also talked about wanting to help future victims of violence and their community as an incentive for participating in research, "At least you're being open to giving you suggestions and making things better for, not, maybe not yourself but somebody else later on down the line." One participant said he wanted to help researchers gather information with the hopes that more studies will develop from the new information. Seven (44%) cross-sectional study participants talked about the societal benefits that can come from someone participating in research, and six (38%) thought researchers should mention the potential societal benefits during the recruitment process to help people decide on whether or not to participate.

> I would say just, um . . . explain it in detail how, by their feedback, can contribute to the research study, to be beneficial for just . . . for society, in general, because it is a research study, and any research study that you do, if it's really successful, the whole . . . the society in general can benefit from it.

Participants' Recommendations for Effective Research Recruitment

Recruitment approach. Two main components to recruitment approach included specific behaviors and the demographic

make-up of researchers with initial patient contact. Participants suggested a mellow approach that begins with a brief introduction. It was important to acknowledge that the potential participant is injured by first inquiring about how they are feeling during the introduction, "I'd start off with flyers, and just giving them the brochure about, what we're about and what are we trying to do . . . Give them . . . and say, make them feel comfortable."

Two (13%) cross-sectional study participants suggested that family may play a role in recruitment. This was apparent in the intervention study, as several participants had their mothers or girlfriends present during the enrollment process and asked these women to come to appointments or help schedule appointments. "So, if you have a family member—somebody's mother, sister, girlfriend, wife—they could persuade them and point them in the right direction and tell them . . . without them have to think about it and evaluate the situation their self."

Two (13%) in the cross-sectional study suggested that the person approaching the potential participant should be someone who appears to come "from the same situation" or be of the same racial background so as to enhance the comfort level.

> Well, I can tell you, for . . . if there was a young Black male sitting in that hospital bed, you bet, you know, your chance would probably be better sending in the youngest Black male person you have workin' for you. 'Cause . . . I know for sure, man, it's all about eyes. And the first thing you see. Before you even open your mouth, I'm seeing you. And if you look different, or if you look, you know, something that I'm not used to seeing, I'm automatically gonna shut down about this much.

The researchers should take great effort to clarify that they are not affiliated with police officers when approaching potential participants in the hospital. Participants suggested providing information about the study to make this distinction.

> So it's all about . . . really . . . showing them that, "We don't mess with the police. That, we don't talk to them or give any information that you give to us TO the police." You know what I'm saying? If you really can prove that to that person . . . you'd get like . . . more people to come in and see you.

Participants also emphasized the importance of making the person feel like this is an "honest situation" and "productive." One talked about offering a study that fits the needs of this population,

> A lot of people need . . . need, need networking and, you know, resources. I mean, if you're just coming to sit there and talk to me, hold my hand, I mean, that's . . . I could call my girlfriend for that . . . a lot of these young cats feel like they don't have too many options. Or, you know, a lot of them don't have a direction yet. So, it all depends on what you're coming to the table with.

Participants said that in the initial contact, researchers should give patients a brochure with a telephone number where they can be reached. Leaving a number on the brochure may help recruit people who plan to follow through on the study. When discussing the content of the recruitment material, some talk about the language that should be included. One participant said the brochure should, "cater more to the actuality of their situation . . . You have to get them inspired first."

Timing. Timing of recruitment was also an important factor given the multiple stressors following an injury: "And reach out when the time is RIGHT. You know what I'm saying? You reach out when the time is right, you'll find people who . . . who are definitely sincere with their injuries and WANT to talk about it." Reasons cited as to why it might be difficult to recruit from the hospital included injury severity, adjusting to new medications, legal issues stemming from the incident, anger, maturity level, and trauma from the injury, "Under them circumstances? You, you are not trying to have that conversation. You know? You're just trying to get taken care of, and . . . you're probably tired, angry, confused . . . you know, just like, want to be left alone."

Eight (50%) cross-sectional participants talked about the emotional sensitivity they felt after their injury that likely would have inhibited them from signing up for a study in the immediate aftermath, with anger specifically mentioned by four participants, "Some people, when you talk to them about a situation, they just get more angry. And if you don't be careful, they'll take it out on YOU!" Participants suggested how long to wait when approaching potential participants that ranged from having time to "rest" and "get comfortable" in the hospital to after "recuperating and settling in" at home.

Discussion

This paper reports on challenges and facilitators to conducting research with Black male victims of community violence. Although some challenges are likely to be common to all populations (e.g., scheduling logistics), the most potent challenges for research participation seen in these studies relate to the inner city street culture in which many of these men are immersed. These barriers, including fear of police, an impression of snitching, and mistrust of institutions, may be explained by what sociologist Elijah Anderson refers to as the "Code of the Street" (Anderson, 1994). He defines this as a set of rules found among inner city Black communities in response to the perceived failure of mainstream institutions to serve their needs, including law enforcement agencies. This Code encourages youth to take control of their own safety by protecting themselves instead of seeking help from, or cooperating with, the police; speaking to the police is viewed as defying the Code. Instead, violent injury may lead victims to carry weapons or seek retaliation to protect themselves (Rich & Gray, 2005). Those who do seek help from law enforcement after an injury may be viewed as weak and unable to defend themselves and vulnerable to future attacks. Black male victims of violence may or may not be actively involved in gangs or street violence. According to Anderson, however, even those who do not have gang affiliations may have shared mistrust of police (Anderson, 1994). This emphasizes the recommendation that research personnel clearly show they are not affiliated with police in printed recruitment materials and via personal reassurance by those conducting recruitment so participants do not feel they are helping police build a case against the perpetrator of their injury by participating in research.

"Stop snitching" is a campaign started among inner city youth meant to discourage cooperation with police regarding investigations into criminal activity (Schorn, 2007). This slogan can be found in music, movies, and clothing. As many study participants mentioned, research participation can be viewed as a form of snitching, particularly if people believe the information is shared with police. Ensuring that participants clearly understand the privacy rules may help alleviate some concern that information can be shared outside of research teams. This may be particularly relevant among participants who use illicit substances, which was an entry criterion for the pilot intervention. However, as seen in this study, the current informed consent

process may not be clear to those who do not typically engage in research.

Concerns about privacy and informed consent are not a novel finding. In one study of African American patients from public and private primary care clinics, participants who reviewed a consent form containing Health Insurance Portability and Accountability Act (HIPAA) authorization were less likely to consider taking part in the study than those who reviewed a consent form with no HIPAA authorization because of mistrust toward the research, research personnel, or research institutions (Dunlop et al., 2007). Working with Institutional Review Boards to make privacy clauses more understandable may help recruit and maintain participants in research studies, particularly with populations that may have issues with trust and privacy. In addition, obtaining a Certificate of Confidentiality from the National Institutes of Health may alleviate some concern of information being shared with law enforcement agents because of the protections it provides. However, researchers need to use terminology that participants understand when explaining the Certificate's protections and be aware of similarities to terminology used by law enforcement.

Previous research has noted the barriers to conducting health care research with African Americans, particularly the mistrust in research institutions (Corbie-Smith et al., 1999; Rajakumar, Thomas, & Musa, 2009; Shavers, Lynch, & Burmeister, 2002). Contributing to this mistrust is knowledge of the Tuskegee study, which may negatively influence people's willingness to participate in research (Shavers, Lynch, & Burmeister, 2002). However, because none of the participants in these studies mentioned this as a reason for mistrust in research and they were not specifically asked about Tuskegee, it is unclear the role it plays among this particular population. Cultural mistrust, stemming from historic and current experiences with racism (Whaley, 2001), may also discourage African American research participation. Studies have shown that African Americans with high levels of cultural mistrust prefer working with Black clinicians and may not be comfortable disclosing personal information to White clinicians (Townes, Chavez-Korell, & Cunningham, 2009). This may relate to sharing personal information to a White research team member. Additionally, among Black male victims of violence, the mistrust in research may also relate to the Code of the Street and fear of getting arrested, reinjured, or killed if a participant discloses too much about the injury circumstances. The importance of cultural competency is often emphasized when conducting research with minority populations (Cooley, Boyd, & Grados, 2004). Awareness of and sensitivity to street culture should be considered when working with inner city populations who may be influenced by the Code of the Street, as well as the influence of cultural mistrust on African Americans' willingness to disclose information. Having a research team member that potential participants feel they can trust may help engage them in research.

Another challenge noted in this study was the difficulty recruiting patients in the early aftermath of the injury. Several factors may influence this, including emotional outcomes experienced after a traumatic event, such as peritraumatic distress and peritraumatic dissociation (Fein et al., 2002; Fein, Kassam-Adams, Vu, & Datner, 2001; Johansen, Wahl, Eilertsen, Hanestad, & Weisaeth, 2006). Logistical issues demonstrated a need for flexibility in scheduling follow-up appointments that related to an injury, such as sleep disturbances, because of physical and emotional outcomes of the injury. The challenge of recruiting victims of violence in the aftermath of the event corroborates findings in other studies working with trauma survivors (Stein et al., 2007). Scott and colleagues tested a study design that considers issues in research with trauma survivors that were found in this study, such as mobility and safety concerns (Scott, Sonis, Creamer, & Dennis, 2006). Similar to Scott's study, researchers should explore different research models for recruiting and randomizing victims of violence in the aftermath of the trauma.

Participants' desire to help others may be an avenue for which studies can increase recruitment. This has been seen among victims of other traumatic experiences (Campbell & Adams, 2009). In our study population, this desire may also stem from the idea of collectivism, the belief that the group or family is at the core of a society (Oyserman, Coon, & Kemmelmeier, 2002), a concept that has been found to be higher among African American populations (Kreuter, Lukwago, Bucholtz, Clark, & Sanders-Thompson, 2003). People with a strong sense of collectivism put their community before themselves and value loyalty, respect, and helping others (Kaniasty & Norris, 2000; Kreuter et al., 2003). Helping Black male victims of violence understand that participating in research can potentially help their community may help increase enrollment into research.

Findings from this study also reveal the high level of financial need among this population. While some may have had this need before their injury, because of the low-income environment from which they came, many were unable to work after their injury because of new physical and emotional disabilities. Communication was a common challenge with study participants, particularly because of the lack of steady income that may otherwise pay for a telephone or the unstable housing that would otherwise provide a steady telephone line. This made it difficult to stay in touch with participants and led to lost participants and rescheduled appointments. Participant availability has been an issue in other studies recruiting inner-city victims of violence because of the frequent mobility found in this population (Zun, Downey, & Rosen, 2006). A study looking at the psychosocial needs of victims of violence found the highest needs to be educational and occupational (Zun & Rosen, 2003). Study incentives catering to the needs of this population, such as money or telephone minutes, may play an important role in getting inner city victims of violence to enroll in studies. Using cellular telephone minutes as an incentive may have dual benefit, both for the participant, as well as for scheduling follow-up appointments.

Through the reflexivity practiced by the research team, investigators tried to be sensitive to what they were asking of participants, especially given the emotional trauma many participants had experienced. It may be appropriate for researchers working with this population to communicate empathy about the burden participating may place on potential subjects. Investigators may also want to emphasize that they are dedicated to preventing violence and assisting victims of violence, and not just there to "profit off participants' misery." They may also comment on the timeline for implementation of changes based on the results so participants do not feel that they provided information with no outcome for themselves or their community.

This study has several limitations. One is the small sample size, which in part may have been influenced by some of the barriers discussed in the results. Information was collected only from those who actually participated in research, thus, information on those who did not participate was gleaned from observation in the intervention component or by conjecture of participants who described how they felt when they were younger. However, a research bias exists toward those who were truly willing to participate in research. In addition,

this data has geographic limitations of an inner-city population in the Northeast, and may not generalize to other locations. The data also may not generalize to Black males who come from more educated, higher socioeconomic, or less urban backgrounds.

In the cross-sectional study, some participants were interviewed by White researchers, which may have influenced their willingness to disclose personal information if they had a high sense of cultural mistrust. The research team included both White and Black members, which may have had inadvertent bias toward the development of codes and themes based on racial or other experiences. To mitigate this, all members of the team reviewed all iterations of the manuscript to ensure that the views represented their own understanding of the issues. The team felt that the differing view points helped articulate, clarify, and interpret findings for each other. In particular, grasping the ramifications of the profound level of mistrust in research and health care systems in study participants spurred a change in consciousness among some team members.

Despite these limitations, this study attempts to provide information that will help in future research with Black male victims of community violence. As a qualitative study, the results provide information on which to base hypotheses for future testing. Future research should test study designs that incorporate the recommendations provided by our sample to see if they lead to greater recruitment, such as timing of recruitment, providing appropriate incentives, and creative communication avenues that may help with scheduling and follow-up. When working with this population in general, researchers should attempt to establish trust by involving peers in the recruitment process and by specifying that they are not affiliated with law enforcement; this may address some of the barriers to research participation. Finally, future research with Black male victims of community violence should be mindful of the cultural context in which these men have grown up and the emotional effects of the traumatic experience. Novel protocols should be adapted to fit within these contexts to maximize their effectiveness.

References

Anderson, E. (1994). The code of the streets. *Atlantic Monthly, 273,* 81–94.

Becker, M. G., Hall, J. S., Ursic, C. M., Jain, S., & Calhoun, D. (2004). Caught in the Crossfire: The effects of a peer-based intervention program for violently injured youth. *Journal of Adolescent Health, 34,* 177–183.

Campbell, R., & Adams, A. E. (2009). Why do rape survivors volunteer for face-to-face interviews? A meta-study of victims' reasons for and concerns about research participation. *Journal of Interpersonal Violence, 24,* 395–405.

Cooley, M., Boyd, R., & Grados, J. (2004). Feasibility of an anxiety prevention intervention for community violence exposed African-American children. *The Journal of Primary Prevention, 25,* 105–123.

Cooper, C., Eslinger, D., Nash, D., al-Zawahri, J., & Stolley, P. (2000). Repeat victims of violence: Report of a large concurrent case-control study. *Archives of Surgery, 135,* 837–843.

Cooper, C., Eslinger, D. M., & Stolley, P. D. (2006). Hospital-based violence intervention programs work. *Journal of Trauma-Injury Infection & Critical Care, 61,* 534–537; discussion 537–540.

Corbie-Smith, G., Thomas, S., & St. George, D. (2002). Distrust, race, and research. *Archives of Internal Medicine, 162,* 2458–2463.

Corbie-Smith, G., Thomas, S., Williams, M., & Moody-Ayers, S. (1999). Attitudes and beliefs of African Americans towards participation in medical research. *Journal of General Internal Medicine, 13,* 537–546.

Dunlop, A., Graham, T., Leroy, Z., Glanz, K., & Dunlop, B. (2007). The impact of HIPAA authorization on willingness to participate in clinical research. *Annals of Epidemiology, 17,* 899–905.

Fein, J. A., Kassam-Adams, N., Gavin, M., Huang, R., Blanchard, D., & Datner, E. M. (2002). Persistence of posttraumatic stress in violently injured youth seen in the emergency department. *Archives of Pediatric & Adolescent Medicine, 156,* 836–840.

Fein, J. A., Kassam-Adams, N., Vu, T., & Datner, E. M. (2001). Emergency department evaluation of acute stress disorder symptoms in violently injured youths. *Annals of Emergency Medicine, 38,* 391–396.

Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory.* Chicago: Aldine.

Johansen, V. A., Wahl, A. K., Eilertsen, D. E., Hanestad, B. R., & Weisaeth, L. (2006). Acute psychological reactions in assault victims of non-domestic violence: Peritraumatic dissociation, post-traumatic stress disorder, anxiety and depression. *Nordic Journal of Psychiatry, 60,* 452–462.

Kaniasty, K., & Norris, F. H. (2000). Help-seeking comfort and receiving social support: The role of ethnicity and context of need. *American Journal of Community Psychology, 28,* 545–581.

Kreuter, M. W., Lukwago, S. N., Bucholtz, R. D., Clark, E. M., & Sanders-Thompson, V. (2003). Achieving cultural appropriateness in health promotion programs: Targeted and tailored approaches. *Health Education and Behavior, 30,* 133–146.

Maruta, T., Swanson, D. W., & Finlayson, R. E. (1979). Drug abuse and dependency in patients with chronic pain. *Mayo Clinic Proceedings, 54,* 241–244.

McCallum, J. M., Arekere, D. M., Green, B. L., Katz, R. V., & Rivers, B. M. (2006). Awareness and knowledge of the U.S. Public Health Service syphilis study at Tuskegee: Implications for biomedical research. *Journal of Health Care for the Poor and Underserved, 17,* 716–733.

National Center for Injury Prevention and Control. (2005). *Leading causes of death reports.* Atlanta, GA: Center for Disease Control and Prevention.

Oyserman, D., Coon, H. M., & Kemmelmeier, M. (2002). Rethinking individualism and collectivism: Evaluation of theoretical assumptions and meta-analyses. *Psychological Bulletin, 128,* 3–72.

Rajakumar, K., Thomas, S., & Musa, D. (2009). Racial differences in parents' distrust of medicine and research. *Archives of Pediatrics and Adolescent Medicine, 163,* 108–114.

Rich, J. A., & Grey, C. M. (2005). Pathways to recurrent trauma among young Black men: Traumatic stress, substance use, and the "code of the street". *American Journal of Public Health, 95,* 816–824.

Rose, S., Brewin, C. R., Andrews, B., & Kirk, M. (1999). A randomized controlled trial of individual psychological debriefing for victims of violent crime. *Psychological Medicine, 29,* 793–799.

Schorn, D. (2007). *Stop snitchin': Rapper Cam'ron: Snitching hurts his business, "Code Of Ethics". 60 Minutes*. Retrieved from http://www.cbsnews.com/stories/2007/04/19/60minutes/main2704565.shtml

Scott, C. K., Sonis, J., Creamer, M., & Dennis, M. L. (2006). Maximizing follow-up in longitudinal studies of traumatized populations. *Journal of Traumatic Stress, 19,* 757–769.

Shavers, V. L., Lynch, C. F., & Burmeister, L. F. (2002). Racial differences in factors that influence the willingness to participate in medical research studies. *Annals of Epidemiology, 12,* 248–256.

Stein, M. B., Kerridge, C., Dimsdale, J. E., & Hoyt, D. B. (2007). Pharmacotherapy to prevent PTSD: Results from a randomized controlled proof-of-concept trial in physically injured patients. *Journal of Traumatic Stress, 20,* 923–932.

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques.* Newbury Park, CA: SAGE Publications.

Townes, D. L., Chavez-Korell, S., & Cunningham, N. (2009). Reexamining the relationships between racial identity, cultural mistrust, help-seeking attitudes, and preference for a black counselor. *Journal of Counseling Psychology, 56,* 330–336.

Warren, C. A. B., & Karner, T. X. (2005). *Discovering qualitative methods: Field research, interviews and analysis.* Los Angeles: Roxbury Publishing Company.

Whaley, A. L. (2001). Cultural mistrust: An important psychological construct for diagnosis and treatment of African Americans. *Professional Psychology: Research and Practice, 32,* 555–562.

Zun, L. S., Downey, L., & Rosen, J. (2006). The effectiveness of an ED-based violence prevention program. *American Journal of Emergency Medicine, 24,* 8–13.

Zun, L. S., & Rosen, J. M. (2003). Psychosocial needs of young persons who are victims of interpersonal violence. *Pediatric Emergency Care, 19,* 15–19.

Received March 6, 2009
Revision received January 25, 2010
Accepted January 25, 2010 ■

HSR Health Services Research

DOI: 10.1111/j.1475-6773.2009.01068.x
RESEARCH ARTICLE

Comparison of Health Outcomes for Male Seniors in the Veterans Health Administration and Medicare Advantage Plans

Alfredo J. Selim, Dan Berlowitz, Lewis E. Kazis, William Rogers, Steven M. Wright, Shirley X. Qian, James A. Rothendler, Avron Spiro III, Donald Miller, Bernardo J. Selim, and Benjamin G. Fincke

Objectives. To compare the Veterans Health Administration (VHA) with the Medicare Advantage (MA) plans with regard to health outcomes.
Data Sources. The Medicare Health Outcome Survey, the 1999 Large Health Survey of Veteran Enrollees, and the Ambulatory Care Survey of Healthcare Experiences of Patients (Fiscal Years 2002 and 2003).
Study Design. A retrospective study.
Extraction Methods. Men 65+ receiving care in MA (N= 198,421) or in VHA (N= 360,316). We compared the risk-adjusted probability of being alive with the same or better physical (PCS) and mental (MCS) health at 2-years follow-up. We computed hazard ratio (HR) for 2-year mortality.
Principal Findings. Veterans had a higher adjusted probability of being alive with the same or better PCS compared with MA participants (VHA 69.2 versus MA 63.6 percent, $p<.001$). VHA patients had a higher adjusted probability than MA patients of being alive with the same or better MCS (76.1 versus 69.6 percent, $p<.001$). The HRs for mortality in the MA were higher than in the VHA (HR, 1.26 [95 percent CI 1.23–1.29]).
Conclusions. Our findings indicate that the VHA has better patient outcomes than the private managed care plans in Medicare. The VHA's performance offers encouragement that the public sector can both finance and provide exemplary health care.

Key Words. Health outcomes, system comparison, quality of care

The Veterans Health Administration (VHA) is one of the largest integrated health care systems in the United States and provides health care to an aging veteran population (Ohldin et al. 2002). In 1995, VHA launched a major

health care transformation by adopting managed care principles (Flynn, McGlynn, and Young 1997). It included the use of information technology, measurement and reporting of performance, and integration of services and realigned payment policies (Kizer, Demakis, and Feussner 2000). By 2000, the VHA performance as measured by processes of care was better than that in the Medicare fee-for-service (FFS) program and in commercial managed care plans (Jha et al. 2003; Asch et al. 2004). However, whether this improvement has extended beyond processes of care and into patient outcomes has not yet been fully examined.

The few studies that have compared patient outcomes in the VHA with other health care systems used data collected in the early stages of the VHA transformation. The evidence on health outcomes for the VHA is mixed. For example, using data from 1993 to 1996, Rosenthal, Kaboli, and Barnett (2003) found that the VHA had higher mortality rates after coronary artery bypass grafting when compared with private sector hospitals serving the same health care market. Landrum et al. (2004) found that elderly male veterans hospitalized for acute myocardial infarction in a VHA facility between October 1, 1996, and September 30, 1999, had higher mortality when

---

Address correspondence to Alfredo J. Selim, M.D., M.P.H., Center for Health Quality, Outcomes & Economic Research, Edith Nourse Rogers Memorial Hospital (152), Building 70, 200 Springs Road, Bedford, MA 01730; e-mail: alfredo.selim2@va.gov. Alfredo J. Selim, M.D., M.P.H., is also Physician with the Emergency Services, Boston VA Health Care System, West Roxbury campus, West Roxbury, MA, and also Assistant Professor of Medicine with the Boston University School of Medicine, Boston, MA. Dan Berlowitz, M.D. M.P.H., is with the Center for Health Quality, Outcomes & Economic Research, Edith Nourse Rogers Memorial Hospital, Bedford, MA. Lewis E. Kazis, Sc.D., Senior Scientist, Shirley X. Qian M.S., SAS Programmer, James A. Rothendler, M.D., Medical Officer, Avron Spiro III, Ph.D., Senior Research Scientist, and Donald Miller, Sc.D., Senior Research Scientist, are with the Department of Veteran Affairs Medical Center, Center for Health Quality, Outcomes and Economic Research, Edith Nourse Rogers Memorial Hospital, Bedford, MA. Lewis E. Kazis, Sc.D., Associate Professor of Health Services, James A. Rothendler, M.D., M.P.H, Assistant Professor of Health Policy and Management, Avron Spiro III, Ph.D, and Donald Miller, Sc.D, Associate Professor of Epidemiology, are also with the Boston University School of Public Health, Boston, MA. William Rogers, Ph.D., is Senior Statistician with The Health Institute, New England Medical Center, Tufts-New England Medical Center, Boston, MA. Steven M. Wright, Ph.D., is Director of Epidemiology for the Office of Quality and Performance, Office of Quality and Performance, U.S. Department of Veterans Affairs, Washington, DC. Bernardo J. Selim, M.D., is Physician with the Yale-New Haven Hospital, New Haven, CT. Benjamin G. Fincke, M.D., is Physician with the Department of Veterans Affairs Medical Center, Patient Care Services, Medical Service, Center for Health Quality, Outcomes & Economic Research, Edith Nourse Rogers Memorial Hospital, Bedford, MA, and also Associate Professor of Medicine with the Boston University School of Medicine, Boston, MA.

compared with FFS Medicare patients. Studies that used data collected in the late 1990s found differences in risk-adjusted mortality that favored the VHA when compared with Medicare programs and private sector hospitals (Petersen et al. 2000; Selim et al. 2006; Vaughan-Sarrazin, Wakefield, and Rosenthal 2007). These contrasting findings leave unresolved questions regarding how outcomes in the VHA compare to that in other systems. The answer is likely to depend on which systems of care are studied, whether one compares overall outcomes for patients or outcomes regarding particular health conditions, as well as the specific metrics that are used.

In a prior study that used data collected between 1998 and 2000, we compared the VHA and the Medicare Advantage (MA) plans (Selim et al. 2007). This comparison was relevant because MA is a Medicare program that provides comprehensive health services to 4.6 million patients (12 percent of the Medicare population) through contracted private managed care organizations (MCOs) throughout the United States (Centers for Medicare and Medicare Services, HHS 2005). The MCOs are heterogeneous and we refer to them collectively as MA. We found that the adjusted probabilities of being alive with the same or better physical health at 2-years follow-up were comparable between the VHA and the MA (63.6 and 64.4 percent, respectively). The adjusted probability of being alive with the same or better mental health at 2-years follow-up in the VHA was significantly higher than in the MA; however, the magnitude of the difference was small (71.8 versus 70.1 percent, respectively). This cross-system comparison study was made at an early stage of the VHA transformation and was based on a limited number of patients. Given the continuing improvements in the VHA in the management of patients with chronic conditions (Asch et al. 2004; Kerr et al. 2004), it is possible that the results would be different with the use of contemporary health outcome databases (Higashi et al. 2005). A study with a larger number of patients might also reveal differences in outcomes for certain vulnerable subpopulations that may be more susceptible to variations in quality of care, such as older patients, patients with multiple medical problems, or racial/ethnic minorities.

In the present study, we sought to verify and extend the cross-system comparisons between the VHA and MA using more contemporary health outcome data and a larger number of patients. We gave special consideration to particular subpopulations that may be at the highest risk of poor health outcomes and in which the VHA has placed a special emphasis on improving care (Asch et al. 2006).

METHODS

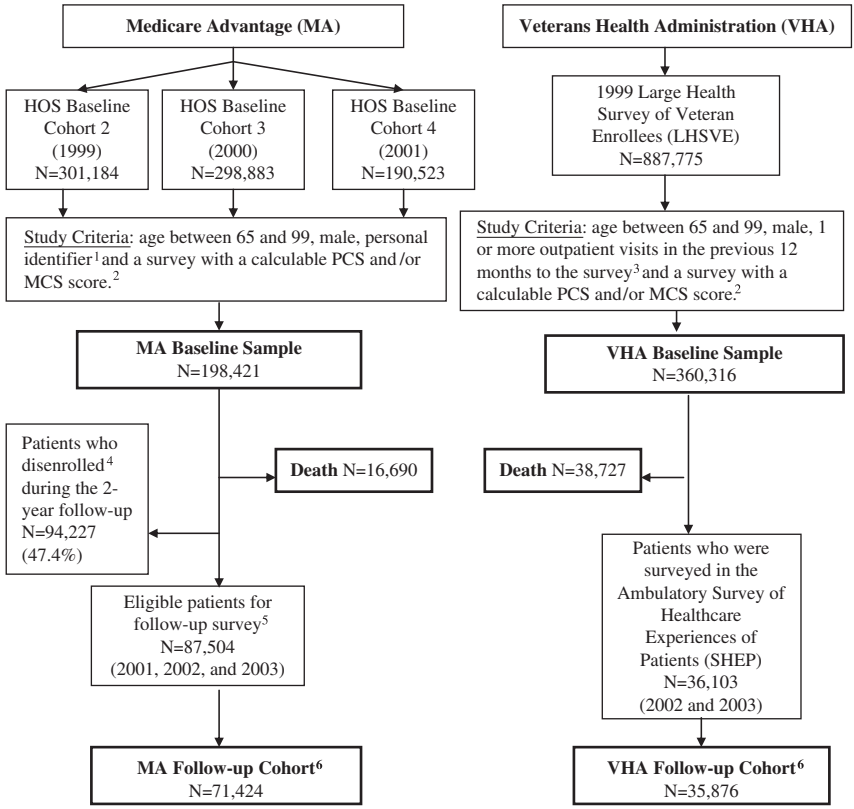
Study Population

The study population from each health care system was restricted to a comparable subset (Figure 1). The MA population was from three cohorts of the Medicare Health Outcomes Survey (HOS): cohorts 2 (1999–2001), 3 (2000–2002), and 4 (2001–2003). The Medicare HOS randomly selected a cohort of 1,000 beneficiaries who were continuously enrolled for at least 6 months from each of the Medicare MCOs (HEDIS® 2003). With the exception of a few specific contract types, all Medicare MCOs participated. The response rates were 66.5, 71, and 68.4 percent at baseline and 84, 77.1, and 78.4 percent at follow-up for cohorts 2, 3, and 4, respectively. The MA population was limited to those beneficiaries aged 65–99 who had person-level identifiers needed to link the HOS data to other databases. Given the disproportionate male representation in the VHA cohort (97.9 percent VHA male patients versus 51.6 percent MA male patients), both MA and VHA analyses were limited to male patients.

The MA population began with a baseline sample of 198,421 male seniors who had filled out the HOS survey and who had enough data to calculate physical (PCS) and mental (MCS) health scores using a previously validated modified regression estimation (MRE) algorithm for missing data (Centers of Medicaid and Medicare Services 2007a). Among the 198,421 Medicare beneficiaries, 16,690 (8.4 percent) died during a 2-year follow-up period. Only Medicare beneficiaries who remained in Medicare managed care ($N = 87{,}504$) were surveyed at 2 years. Among the 87,504 patients who were surveyed at follow-up, 71,424 had enough SF-36® data to calculate the PCS and MCS scores using the MRE algorithm.

The VHA population was derived from the 1999 Large Health Survey of Veteran Enrollees (LHSVE), a cross-sectional health survey administration (Perlin et al. 2000). The 1999 LHSVE randomly selected a total of 1.5 million veterans from the enrollment file of 3,613,877 individuals. The enrollment file consisted of all veterans registered with the VHA for medical care. The baseline survey administration took place between July 1999 and January 2000. The response rate was 63.1 percent. For the purpose of this study, the analyses were limited to male veterans aged 65–99 who had 1 or more outpatient visits in the previous 12 months and had enough data to calculate PCS and MCS scores using the MRE algorithm (Kazis et al. 2004b). Thus, the VHA baseline sample consisted of 360,316 elderly male veterans. Among the 360,316 veterans, 38,727 (10.6 percent) died during a 2-year follow-up

Figure 1: Medicare Advantage and Veterans Health Administration Populations



[1] Primary Medicare beneficiaries have a personal identifier (SSN) to link them with non-HOS databases in the analysis.

[2] Patients that responded a health survey (complete, partial complete) and had enough data to score the physical (PCS) or mental (MCS) component summary scores using the modified regression estimates (MRE) algorithms.

[3] This is a sampling strategy that the VHA applies for health care evaluations.

[4] Patients disenrolled from managed care organizations because their plan no longer participated in Medicare managed care or the beneficiaries themselves were no longer enrolled in the same plan.

[5] Patients were eligible for follow-up survey administrations only if they stayed in the same managed care organization in the two years follow-up period.

[6] Patients that responded a health survey (complete, partial complete) and had enough data to score the physical (PCS) or mental (MCS) component summary scores using the modified regression estimates (MRE) algorithms.

period. The VHA follow-up sample was derived from the Ambulatory Care Survey of Healthcare Experiences of Patients (SHEP), which is a patient-centered initiative by the Office of Quality and Performance to assess

satisfaction, functional status, and health behavior information from veterans who obtain care in the VHA (Wright et al. 2006). The SHEP had a response rate of 70.3 percent. Among the 360,316 patients from our baseline cohort, we found that 36,103 veterans were included in SHEP during the fiscal years 2002 and 2003. Only three patients did not respond to the survey. Of the 36,100 veterans, 35,876 had enough data to calculate PCS or MCS scores using the MRE algorithm (Centers of Medicaid and Medicare Services 2007b).

Tables SA1 and SA2 show the comparisons of baseline patient characteristics between the beneficiaries who stayed (enrollees) and those who did not stay (disenrollees) in MA at 2 years and veterans who were surveyed in SHEP and those who were not surveyed. Although the patient characteristics differences were statistically significant, the magnitude of the differences was very small, thus indicating that MA enrollees and disenrollees were similar as were veterans who were surveyed and not surveyed in SHEP.

Survey Methods

The Medicare HOS used the Medical Outcomes Study (MOS) Short Form-36 (SF-36®) Health Survey at baseline and at follow-up to measure health status (Gandek et al. 2004). For the VHA, the Veterans RAND 36-Item Health Survey (VR-36) was used for baseline health status, and the Veterans RAND 12-Item Health Survey (VR-12) was used for follow-up (Kazis et al. 2004a). They differ from the SF-36® in the use of five-point response choices for the role limitations due to physical problems and the role limitations due to emotional problems. There are validated conversion formulas that allow for comparisons of VR-36 and VR-12 scores to those from the SF-36® (Kazis et al. 2004b). We summarized the health surveys into physical (PCS) and mental component (MCS) scales by applying a linear *t*-score transformation with a mean of 50 and a standard deviation of 10 based on the 1990 U.S. population norms (Ware, Kosinski, and Keller 1994, 1995). We accounted for the differences in follow-up sampling strategies between the VHA and the MA by applying sampling weights to the PCS and MCS scores.

Outcome Measures

Based on the concept that patient health lies on a continuum from well-being to death, we used three outcomes: (1) the probability of being alive with the same or better (than would be expected by chance) PCS at 2 years, (2) the probability of being alive with the same or better (than would be expected by chance) MCS at 2 years, and (3) 2-year mortality. A 2-year follow-up was

selected because studies have shown significant changes in health status for this duration (Lorig et al. 2001). The composite outcome measures of being alive with "the same or better" PCS (or MCS) at follow-up were selected because they reflect the health care goal of maintaining or improving the physical and mental health status of elderly patients (Cohen et al. 2002). The cut-off points for the operational definition of change were "two standard errors of the measurement" (SEM) for a single score or 1.414 standard errors of change. We have applied this cut-off methodology for the outcome metrics because changes of this magnitude have been shown to be clinically relevant (Ware, Kosinski, and Keller 1994) and have been validated in previous studies (Ware et al. 1996). Mortality, our other outcome, is a measure that is particularly relevant to elderly patients and might reflect potentially poor quality of care (Reason 2000). We used the Death Master File from the Social Security Administration to ascertain vital status (Schall et al. 2001).

Case Mix Variables for Risk-Adjustment

Based on prior work (Selim et al. 2007), we used three domains of risk: sociodemographics, comorbidities, and baseline health status. Since the risk for health outcomes differs by demographic subgroup, we selected the following sociodemographic variables: age, race/ethnicity, marital status, education, and income (Williams 1996). Because substantial variations in health status among patients with different diagnoses exist, we selected a group of conditions that are commonly encountered in clinic visits and are known to be major indicators of health status (Bayliss et al. 2004). These are self-reported diagnoses of coronary artery disease (CAD)/myocardial infarction, angina, congestive heart failure (CHF), stroke, hypertension, diabetes, chronic obstructive pulmonary disease (COPD) and asthma, and cancer (other than skin cancer). We included baseline PCS and MCS scores in the risk-adjustment models for being alive because they are important predictors of survival (Dorr et al. 2006). The baseline health scores were not included in the change in PCS (or MCS) models because their coefficients are influenced by the baseline score measurement error and intertemporal correlation. All the case mix variables in our analyses correspond to characteristics of individual patients obtained in their baseline surveys.

Statistical Analysis

In order to examine the differences in patient characteristics between MA and VHA, we compared them in terms of sociodemographics, comorbid

conditions, and PCS and MCS scores. Because social circumstances are known to affect medical outcomes, a social disadvantage index was generated and included minority, unmarried, <12 years of education, and income less than U.S.$20,000 (Anand et al. 2006). To compare the differences in the two study groups at baseline, we used chi-square tests for categorical variables and *t*-tests for continuous variables.

To compare change in health status between the MA and VHA, we applied a previously validated methodology (Selim et al. 2007). We first calculated the change in PCS (or MCS) points per unit of time for each health care system. We took the difference between the PCS (or MCS) baseline and follow-up scores and divided it by the median difference, in years, between the baseline and follow-up surveys (in the MA and the VHA these numbers were 2.09 and 3.29 years, respectively). The reason for using the median time was that late surveys tend to be different (in the HOS they tend to be sicker respondents) and follow-up protocols vary. This gave us the change in points per year, which we multiplied by 2 to get the change in points per 2 years. If the change in points has mean μ_D and standard deviation σ_D and the percentage the same or better has mean μ_P and standard deviation σ_P, then the conversion formula to change an average point difference D into a percentage change P is

$$\frac{P-\mu_P}{\sigma_P}=\frac{D-\mu_D}{\sigma_D}$$

The terms were rearranged to a more familiar linear equation form, $P=aD+b$, where P is the probability of PCS (or MCS) the same or better; D is the mean change in PCS (or MCS) points per 2 years for each health care system, which was adjusted using least-squares means to control for sociodemographics and comorbidities; and a is the slope; and b is the intercept. We computed the "slope" and the "intercept" values using harmonic regressions because neither x (change in PCS [or MCS] points per 2 years) nor y (probability of PCS [or MCS] the same or better) is thought of as the dependent variable (Shalabh 2001). It is the harmonic mixture of regressing y on x and x on y. We applied the harmonic regressions to the HOS cohort because we needed to look to some unit of aggregation above the individual respondent in order to do the conversion and the HOS provides a large dataset naturally organized into managed care plans. For example, if PCS goes down by 1 point per year on the average, the converted probability of PCS the same or better is $5.03(-2)+80.61=70.55$ percent. We then combined the probability of PCS (or MCS) the same or better and the probability of being alive (or 1-probability of

death) using the following formulas (Diehr et al. 2003):

$$\text{Probability of being alive with PCS the same or better} = (1 - \text{Prob death}) \times (\text{Prob PCS the same or better})$$

$$\text{Probability of being alive with MCS the same or better} = (1 - \text{Prob death}) \times (\text{Prob MCS the same or better})$$

We computed the probability of being alive separately from the probability of having the same or better PCS (or MCS) at 2 years for two reasons. First, the population on which the vital status can be assessed with reasonable case mix controls consists of all persons who completed the baseline survey. This is a different sample from those who were followed up with survey administrations. Second, the vital status came from a source that was independent from the survey administrations, and therefore did not share any correlated error with baseline assessments. Thus, we used the full baseline sample of 360,316 VHA and 198,421 MA patients for the calculation of the probability of being alive, which was adjusted for sociodemographics (age, race/ethnicity, marital status, education, and income), comorbidities, and baseline health status (PCS and MCS scores). For the calculation of the probability of having the same or better PCS (or MCS) at 2 years, we used the follow-up sample of 35,876 VHA and 71,424 MA patients. We compared the adjusted probability of being alive with the same or better PCS (or MCS) at 2 years between the VHA and MA using tests of significance based on the standard error of the difference:

$$SE(p_1 - p_2) = \sqrt{\left(\frac{p_1(100 - p_1)}{n_1} + \frac{p_2(100 - p_2)}{n_2}\right)}$$

We also conducted additional analysis applying propensity score matching to examine patients who look alike in all observed characteristics.

We used Cox regression models to estimate hazard ratios (HRs) of dying with 95 percent confidence intervals (CI) for the MA compared with the VHA patients. We calculated the adjusted 2-year mortality rates for each health care system as its observed mortality rate divided by its expected mortality rate, multiplied by the mean of the observed mortality rate for all study patients across both systems.

We also examined subpopulations defined by specific patient characteristics since there might be especially large differences between the VHA and the MA, including age (75 and older), ethnicity/race (whites, African Americans, and Hispanics), and selected chronic conditions (diabetes,

hypertension, CAD/myocardial infarction, and CHF). The stratification variables were not included within the subgroup analyses as risk adjustors. The test of significance was adjusted for multiple comparisons using a Bonferroni correction. We divided 0.05 by the 20 group comparisons of the probability of being alive with the same or better PCS (or MCS) at 2-years follow-up between the MA and VHA, resulting in a significant level of 0.0025 as the cut-off.

RESULTS

Table 1 shows the sociodemographic characteristics of patients in MA and VHA. Mean age was 74.0 (SD ± 6) for the MA patients, while it was 73.7 (SD ± 5) for the VHA patients. The VHA patients, in comparison with MA plan patients, were significantly more likely to be nonwhites (17.8 versus 11.0 percent), were less likely to be married (68.9 versus 77.4 percent), were more

Table 1: Sociodemographic Characteristics of Medicare Advantage (MA) and Veterans Health Administration (VHA) Patients

	Medicare Advantage (N= 198,421)	*Veterans Health Administration (N= 360,316)*
Age, years (SD)	74.0 (± 6)	73.7 (± 5)
Race/ethnicity		
Whites	89.0%	82.2%
African Americans	6.4%	8.7%
Hispanics	1.8%	4.8%
Other	2.8%	4.3%
Marital status (married)	77.4%	68.9%
Education (<12 years)	30.7%	38.9%
Income (<U.S.$20,000)	41.9%	65.6%
Social disadvantage index*		
0	38.1%	17.5%
1	29.6%	33.5%
2	22.0%	31.2%
3	8.7%	14.6%
4	1.6%	3.3%

Note. All comparisons between MA and VHA were significant at <.0001

*The social disadvantage index includes minority, unmarried, <12 years of education, and income less than U.S.$20,000. A higher score indicates greater disadvantage.

likely to have <12 years of education (38.9 versus 30.7 percent), and were more likely to earn an income of less than U.S.$20,000 (65.6 versus 41.9 percent). Overall, the VHA had a higher level of social disadvantage than MA.

Table 2 shows the clinical features of patients in MA and VHA at baseline. The VHA patients had a significantly higher disease burden than the MA patients. The VHA patients were more likely to have four or more chronic medical conditions when compared with MA patients (23.5 versus 9.5 percent, respectively). VHA patients, in comparison with MA plan patients, were significantly more likely to have hypertension (65.7 versus 52.2 percent), coronary artery disease/myocardial infarction (28.3 versus 15.7 percent), diabetes (28.2 versus 19.8 percent), COPD/asthma (25.8 versus 13.5 percent), cancer (19.7 versus 15.1 percent), and stroke (15.3 versus 9.3 percent). VHA patients, in comparison with MA plan patients, had significantly worse physical health (PCS scores, 35.7 versus 43.3) and mental health (MCS scores, 45.2 versus 51.9) at baseline.

Table 2: Clinical Features of Medicare Advantage (MA) and Veterans Health Administration (VHA) Patients

	Medicare Advantage (N= 198,421)	*Veterans Health Administration (N= 360,316)*
Number of comorbidities (range 0–8)		
0	25.9%	12.6%
1	32.4%	25.1%
2	21.2%	23.0%
3	11.2%	15.8%
4 or more	9.5%	23.5%
Diabetes	19.8%	28.2%
Hypertension	52.2%	65.7%
Angina	20.7%	34.5%
Coronary artery disease/myocardial infarction	15.7%	28.3%
Congestive heart failure	8.7%	24.9%
Stroke	9.3%	15.3%
Chronic obstructive pulmonary disease/asthma	13.5%	25.8%
Cancer	15.1%	19.7%
Baseline physical health-PCS, points (SD)*	43.3 (± 11)	35.7 (± 10)
Baseline mental health MCS, points (SD)*	51.9 (± 10)	45.2 (± 13)

Note. All comparisons between MA and VHA were significant at <.0001.

*A lower number is indicative of poor health status for MCS and PCS.

Table 3 shows that the adjusted probability of being alive with the same or better PCS at 2-years follow-up was significantly higher in the VHA when compared with the MA (69.2 versus 63.6 percent, respectively). The adjusted probability of being alive with the same or better MCS at 2-years follow-up in the VHA was also significantly higher than in the MA (76.1 versus 69.6 percent, respectively). The propensity score matching analysis showed comparable results (the probability of being alive with the same or better PCS was 69.3 versus 63.5 percent and the probability of being alive with the same or better MCS was 75.9 versus 69.6 percent for the VHA and MA, respectively). The adjusted 2-year mortality rates were 11.8 and 9.9 percent for the MA, and VHA, respectively, with a significantly higher HR for mortality in the MA compared with the VHA (HR, 1.26 [95 percent CI 1.23–1.29]).

Table 3: Change in Health Status in the Medicare Advantage (MA) and the Veterans Health Administration (VHA)

Health Care System	*MA*	*VHA*
Baseline sample	198,421	360,316
Follow-up cohort	71,424	35,876
Adjusted probability of being alive at 2 years*	88.1%	90.1%
Adjusted probability of the same or better PCS at 2 years†	72.9%	76.8%
Adjusted probability of the same or better MCS at 2 years†	79.0%	84.4%
Adjusted probability of being alive* with the same or better PCS† at 2 years‡	63.6%	69.2%
Adjusted probability of being alive* with the same or better MCS† at 2 years‡	69.6%	76.1%
Propensity score matching		
Probability of being alive with the same or better PCS at 2 years	63.5%	69.3%
Probability of being alive with the same or better MCS at 2 years	69.6%	75.9%
Mortality		
2-years adjusted mortality rates*	11.8%	9.9%
Odds ratio (95% confidence interval)	1.26 (1.23–1.29)	1

*The adjustment variables for survival/mortality included sociodemographics (age, race/ethnicity, marital status, education, and income), comorbidities (hypertension, diabetes, coronary artery disease/myocardial infarction, congestive heart failure, stroke, chronic obstructive pulmonary disease [COPD] and asthma, and cancer [other than skin cancer]), and baseline health status (PCS and MCS scores).

†The adjustment variables for PCS (or MCS) the same or better included sociodemographics (age, race/ethnicity, marital status, education, and income) and comorbidities (hypertension, diabetes, coronary artery disease/myocardial infarction, congestive heart failure, stroke, chronic obstructive pulmonary disease [COPD] and asthma, and cancer [other than skin cancer]).

‡All comparisons between MA and VHA were significant at p-value $< .001$.

We found that across the board, subgroups of vulnerable male patients had better outcomes in the VHA compared with those in MA (Table 4). The adjusted probabilities of being alive with the same or better PCS (or MCS) at 2 years in the VHA were also significantly higher than those in the MA across the very old patients (75 and older) as well as the race/ethnicity groups, including whites, African Americans, and Hispanics. Similar findings are also seen across patients with selected chronic conditions, including diabetes, hypertension, CAD/myocardial infarction, and CHF. The magnitude of the differences in the percentage of patients who were either "alive with the same or better PCS" or "alive with the same or better MCS" at 2-years follow-up ranged from 3 to 10 percent higher in the VHA when compared with the MA. The HRs for mortality in the MA were significantly higher than those in the VHA across all subgroups.

DISCUSSION

After adjusting for the higher prevalence of chronic disease and worse self-reported baseline health status in the VHA, we found significant differences in 2-year health outcomes that favor the VHA when compared with the MA. This was true for the average elderly male patient cared for in the VHA as well as for vulnerable subpopulations. This study, based on data collected on males between 1999 and 2003, extends our previous work using data collected between 1998 and 2000 (Selim et al. 2007).

One might argue that the "execution" of the VHA transformation may have contributed to the better health outcomes found in our study. Although each organizational transformation is unique, VHA's experiences offer a number of lessons for future transformations. The transformation framework included the creation of a vision for the future, the adoption of a new organization structure, careful planning and monitoring of performance to achieve system-wide coordination and accountability and the modification in rules and regulations for access to care (Young 2006). The VHA transitioned from a tertiary/specialty and inpatient-based care system delivering care in a traditional professional model into a system focused on ambulatory care. It developed integrated health care networks that have a collective goal of delivering services to a defined population in a coordinated and collaborative manner that maximizes the health care value of the service (Kizer, Demakis, and Feussner 2000). The VHA enhanced equity of access among ethnic minority veterans and mental health care services and implemented a

Table 4: Change in Health Status in the Medicare Advantage (MA) and the Veterans Health Administration (VHA) by Subpopulations

Subpopulations	*Health Care System*	*Baseline Sample*	*Follow-up Cohort*	*Adjusted Probability of Being Alive* with the Same or Better PCS† at 2 Years‡ (%)*	*Adjusted Probability of Being Alive* with the Same or Better MCS† at 2 Years‡ (%)*	*Adjusted Mortality Rates* (%)*	*Odds Ratio (95% Confidence Interval)*
Age 75 and older	MA	77,405	34,216	68.0	73.7	18.2	1.48 (1.43–1.52)
	VHA	159,159	14,647	72.4	79.1	13.1	1
Whites	MA	176,621	78,147	64.1	69.6	11.9	1.26 (1.23–1.30)
	VHA	294,108	30,622	69.1	76.2	10.0	1
African Americans	MA	12,688	4,981	66.2	68.4	11.9	1.12 (1.03–1.22)
	VHA	31,276	2,248	70.8	73.9	10.3	1
Hispanics	MA	3,557	1,519	65.6	69.8	9.7	1.36 (1.16–1.60)
	VHA	17,012	1,508	71.9	77.0	7.3	1
Hypertension	MA	101,897	44,505	64.5	69.4	12.2	1.24 (1.20–1.28)
	VHA	231,265	23,728	68.8	76.1	10.2	1
Diabetes	MA	38,754	16,229	62.2	66.6	14.9	1.24 (1.19–1.29)
	VHA	98,228	10,148	67.4	73.9	12.4	1
Coronary artery disease	MA	30,336	12,564	62.1	66.0	16.7	1.25 (1.19–1.30)
	VHA	96,927	9,903	66.6	73.9	13.8	1
Congestive heart failure	MA	16,765	6,322	58.4	60.4	23.3	1.28 (1.23–1.34)
	VHA	84,415	7,732	62.8	70.8	19.0	1

*The adjustment variables for survival/mortality included sociodemographics (age, race/ethnicity, marital status, education, and income), comorbidities (hypertension, diabetes, coronary artery disease/myocardial infarction, congestive heart failure, stroke, chronic obstructive pulmonary disease [COPD] and asthma, and cancer [other than skin cancer]), and baseline health status (PCS and MCS scores).

†The adjustment variables for PCS (or MCS) the same or better included sociodemographics (age, race/ethnicity, marital status, education, and income) and comorbidities (hypertension, diabetes, coronary artery disease/myocardial infarction, congestive heart failure, stroke, chronic obstructive pulmonary disease [COPD] and asthma, and cancer [other than skin cancer]).

The stratification variables were not included within the subgroup analyses as risk adjustors.

‡All comparisons between MA and VHA were significant at $p<.0025$ based on Bonferroni correction.

sophisticated electronic health record system that improved quality, efficiency, and costs of medical care (Wooten 2002; Chaudhry et al. 2006; Kressin et al. 2007).

The magnitude of the VHA's improved performance in the subgroup analysis is not better than in the full VHA sample, indicating that process improvements in the VHA are broad and not limited to an area in which special programs exist. Research studies in documenting and evaluating the transformation have found similarly broad effects when examining the progress and diverse impacts of the overall reorganization and several of its specific elements, including the creation of a seamless continuum of care, making superior quality consistent and demonstrating good value (Feussner, Kizer, and Demakis 2000; Berlowitz et al. 2001; Jha et al. 2003). Therefore, the creation of a clear vision of the future and a coherent transformation plan having concrete and concise goals and performance measures are essential to a successful health care system based on managed care principles.

There are a number of limitations in this study that might affect our results. First, the MA respondents started with higher scores than the VHA respondents. The reasoning might be that the scores have more to drop in MA because they are higher to start with and would suggest a model where all groups are progressing toward death at a fixed time. We have found that after correcting for regression to the mean effects, those individuals with lower scores start to decline at the same rate or faster than those with higher scores.

Second, the health survey instruments may be insensitive to further decline, which would bias against documenting worsening health status in patients who are already severely ill ("floor effect"). Against this is the finding of other investigators that over half of the patients with low health status were able to report that their health status subsequently declined further (Bindman, Keane, and Lurie 1990).

Third, the two surveys were conducted using different sampling strategies. The MA sample was based on sampling at the plan level, with follow-up survey data subject to continued plan participation in Medicare and the survey respondent's continuous enrollment in the plan. The baseline VHA sample was a population-based survey of all VHA patients and the follow-up sample were those in SHEP. Figure SA1 indicates that MA baseline cohort and those who were surveyed at follow-up were similar as were VHA baseline cohort and those who were surveyed in SHEP. The time between assessments was also different in the two systems. Therefore, the estimates of changes in health status at 2 years are a linear interpolation of changes.

Fourth, this study examines only male patients. There may be differences in outcomes between men and women within the VHA and MA, and for women, comparisons of outcomes between the VHA and MA may differ from that seen in men. Further research is needed to assess the effect of systems of care on outcomes for women.

Fifth, controlling for sociodemographics and comorbid illnesses explained only a fraction of the variance in the change in PCS and MCS. The pseudo R^2 was 0.0021 for PCS the same or better and 0.0039 for MCS the same or better. The same has been true in other studies (Bayliss et al. 2004). The mortality model had a *c*-statistic (discriminative power) of 0.747, which was equal or superior to values obtained in risk-adjusted mortality models for inpatient populations (Best and Cowper 1994; Schneeweiss et al. 2003). Our risk-adjustment methodology did not control for unobservable variables such as quality of care that could be correlated with the health care systems.

Sixth, our outcome measures combine respondents with good health status at baseline who do not decline over the time-frame with respondents in poor health status at baseline who improve. We broke down the composite outcomes into better and same. The categories of change in PCS were classified as (1) "the same" (or unchanged) between -5.66 points and $+5.66$ points, (2) "better" as $> +5.66$ points, and (3) "worse" as < -5.66 points. The categories of change in MCS were classified as (1) "the same" (or unchanged) between -6.72 points and $+6.72$ points, (2) "better" as $> +6.72$ points, and (3) "worse" as < -6.72 points. The cross-system comparison analysis yielded similar results (Table SA3).

Seventh, MA plans are independent and heterogeneous. In particular, there are those that probably look more like the VHA in terms of integration, use of health information technology, performance monitoring, and financial incentives and others that are probably closer to the unmanaged, FFS population. Further studies are needed to evaluate how differences in the organizational characteristics of variations in MA are associated with variation in and the impact on their performance.

Eighth, our work was limited to the comparison of patient outcomes in different settings of managed care. Some evidence suggests elderly patients with chronic conditions enrolled in Medicare managed care experienced greater declines in physical health than similar persons in Medicare FFS (Shaughnessy, Schlenker, and Hittle 1994; Ware et al. 1996). More recent studies of the elderly have found no differences between Medicare managed care and FFS with respect to functional declines and to mortality (Retchin et al.

1992; Riley 2000; Porell and Miltiades 2001). Further research is needed to better understand how these different systems of care affect outcomes.

In summary, our findings indicate that the VHA has better outcomes for men than MA. The VHA's performance offers encouragement that the public sector can both finance and provide exemplary health care. The VHA's experience provides some general, potentially transferable, and useful policy directions that might benefit other health care systems in the public as well as the private sectors.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: We thank Dr. Samuel C. Haffer from the Medicare HOS Program Office of Research, Development, & Information, Centers for Medicare and Medicaid Services for his thorough review of this manuscript.

Disclosures: The research in this article was supported by the Centers for Medicare and Medicaid Services and the National Committee for Quality Assurance (NCQA) under contract numbers 500-00-0055 with NCQA, Office of Quality and Performance (OQP) of the Department of Veterans Affairs, The Center for Health Quality, Outcomes and Economic Research (CHQOER), Department of Veterans Affairs and Boston University School of Public Health.

Disclaimers: The views expressed in this article are those of the authors and do not necessarily reflect the views of the U.S. Department of Health and Human Services, the Centers for Medicare and Medicaid Services, the National Committee for Quality Assurance, the Department of Veterans Affairs, or Boston University.

The SF-36® Health Survey is a registered trademark of the Medical Outcomes Trust.

REFERENCES

Anand, S. S., F. Razak, A. D. Davis, R. Jacobs, V. Vuskan, K. Teo, and S. Yusuf. 2006. "Social Disadvantage and Cardiovascular Disease: Development of an Index and Analysis of Age, Sex, and Ethnicity Effects." *International Journal of Epidemiology* 35: 1239–45.

Asch, S. M., E. A. McGlynn, M. M. Hogan, R. A. Hayward, P. Shekelle, L. Rubenstein, J. Keesey, J. Adams, and E. A. Kerr. 2004. "Comparison of Quality of Care for

Patients in the Veterans Health Administration and Patients in a National Sample." *Annals of Internal Medicine* 141: 938–45.

———. 2006. "Who Is at Greatest Risk for Receiving Poor-Quality Health Care?" *New England Journal of Medicine* 354: 1147–56.

Bayliss, E. A., M. S. Bayliss, J. E. Ware Jr., and J. F. Steiner. 2004. "Predicting Declines in Physical Function in Persons with Multiple Chronic Medical Conditions: What We Can Learn from the Medical Problem List." *Health Quality of Life Outcomes* 2: 47.

Berlowitz, D. R., G. J. Young, G. H. Brandeis, B. Kader, and J. J. Anderson. 2001. "Health Care Reorganization and Quality of Care: Unintended Effects on Pressure Ulcer Prevention." *Medical Care* 39: 138–46.

Best, W. R., and D. C. Cowper. 1994. "The Ratio of Observed-to-Expected Mortality as a Quality of Care Indicator in Non-Surgical VA Patients." *Medical Care* 32: 390–400.

Bindman, A. B., D. Keane, and N. Lurie. 1990. "Measuring Health Changes among Severely Ill Patients. The Floor Phenomenon." *Medical Care* 28: 1142–52.

Centers for Medicare and Medicare Services, HHS. 2005. "Medicare Program; Establishment of the Medicare Advantage Program; Interpretation. Final Rule; Interpretation." *Federal Register* 70: 13401–2.

Centers of Medicaid and Medicare Services. 2007a. "Imputing the Physical and Mental Summary Scores (PCS and MCS) for the MOS SF-36 and the Veterans SF-36 Health Survey in the Presence of Missing Data" [accessed on July 30, 2007]. Available at http://www.hosonline.org/surveys/hos/download/HOS_Veterans_36_Imputation.pdf

Centers of Medicaid and Medicare Services. 2007b. "Imputing Physical and Mental Summary Scores (PCS and MCS) for the Veterans SF-12 Health Survey in the Context of Missing Data" [accessed on July 30, 2007]. Available at http://www.hosonline.org/surveys/hos/download/HOS_Veterans_12_Imputation. pdf

Chaudhry, B., J. Wang, S. Wu, M. Maglione, W. Mojica, E. Roth, S. C. Morton, and P. G. Shekelle. 2006. "Systematic Review: Impact of Health Information Technology on Quality, Efficiency, and Costs of Medical Care." *Annals of Internal Medicine* 14: 742–52.

Cohen, H. J., J. R. Feussner, N. Weinberger, M. Carnes, R. C. Hamdy, F. Hsieh, C. Phibbs, D. Courtney, K. W. Lyles, C. May, C. McMurtry, L. Pennypacker, D. M. Smith, N. Ainslie, T. Hornick, K. Brodkin, and P. Lavori. 2002. "A Controlled Trial of Inpatient and Outpatient Geriatric Evaluation and Management." *New England Journal of Medicine* 346: 905–12.

Diehr, P., D. L. Patrick, M. B. McDonell, and S. D. Fihn. 2003. "Accounting for Deaths in Longitudinal Studies Using the SF-36: The Performance of the Physical Component Scale of the Short Form 36-Item Health Survey and the PCTD." *Medical Care* 41: 1065–73.

Dorr, D. A., S. S. Jones, L. Burns, S. M. Donnelly, C. P. Brunker, A. Wilcox, and P. D. Clayton. 2006. "Use of Health-Related, Quality-of-Life Metrics to Predict Mortality and Hospitalizations in Community-Dwelling Seniors." *Journal of the American Geriatric Society* 54: 667–73.

Feussner, J. R., K. W. Kizer, and J. G. Demakis. 2000. "The Quality Enhancement Research Initiative: From Evidence to Action." *Medical Care* 38: I1–6.

Flynn, K., G. McGlynn, and G. Young. 1997. "Transferring Managed Care Principles to VA." *Hospital Health Service Administration* 42: 323–38.

Gandek, B., S. J. Sinclair, M. Kosinski, and J. E. Ware Jr. 2004. "Psychometric Evaluation of the SF-36 Health Survey in Medicare Managed Care." *Health Care Finance Review* 25: 5–25.

Jha, A. K., J. B. Perlin, K. W. Kizer, and R. A. Dudley. 2003. "Effect of the Transformation of the Veterans Affairs Health Care System on the Quality of Care." *New England Journal Medicine* 348: 2218–27.

Kazis, L. E., D. R. Miller, K. M. Skinner, A. Lee, X. S. Ren, J. A. Clark, W. H. Rogers, A. Spiro III, A. Selim, M. Liozer, S. M. Payne, D. Manzell, and B. G. Fincke. 2004a. "Patient-Reported Measures of Health: The Veterans Health Study." *Journal of Ambulatory Care Management* 27: 70–83.

———. 2004b. "Improving the Response Choices on the Veterans SF-36 Health Survey Role Functioning Scales: Results from the Veterans Health Study." *Journal of Ambulatory Care Management* 27: 263–80.

Kerr, E. A., R. B. Gerzoff, S. L. Krein, J. V. Selby, J. D. Piette, J. B. Curb, W. H. Curb, W. H. Herman, D. G. Marrero, K. M. Narayan, M. M. Safford, T. Thompson, and C. M. Mangione. 2004. "Diabetes Care Quality in the Veterans Affairs Health Care System and Commercial Managed Care: The TRIAD Study." *Annals of Internal Medicine* 14: 272–81.

Kizer, K. W., J. G. Demakis, and J. R. Feussner. 2000. "Reinventing VA Health Care: Systematizing Quality Improvement and Quality Innovation." *Medical Care* 38 (6, suppl 1): I7–16.

Kressin, N. R., M. E. Glickman, E. D. Peterson, J. Whittle, M. B. Orner, and L. A. Petersen. 2007. "Functional Status Outcomes among White and African-American Cardiac Patients in an Equal Access System." *American Heart Journal* 153: 418–25.

HEDIS®. 2003. *Specifications for the Medicare Health Outcomes Survey*, Vol. 6. Washington, D.C., National Committee for Quality Assurance.

Higashi, T., P. G. Shekelle, J. L. Adams, C. J. Kamberg, C. P. Roth, D. H. Solomon, D. B. Reuben, J. Chiang, C. H. MacLean, J. T. Chang, R. T. Young, D. M. Saliba, and N. S. Wenger. 2005. "Quality of Care Is Associated with Survival in Vulnerable Older Patients." *Annals of Internal Medicine* 143: 274–81.

Landrum, M. B., E. Guadagnoli, R. Zummo, D. Chin, and B. J. McNeil. 2004. "Care Following Acute Myocardial Infarction in the Veterans Administration Medical Centers: A Comparison with Medicare." *Health Services Research* 39: 1773–92.

Lorig, K. R., P. Ritter, A. L. Stewart, D. S. Sobel, B. W. Brown Jr., A. Bandura, V. M. Gonzalez, D. D. Laurent, and H. R. Holman. 2001. "Chronic Disease Self-Management Program: 2-Year Health Status and Health Care Utilization Outcomes." *Medical Care* 39: 1217–23.

Ohldin, A., R. Taylor, A. Stein, and T. Garthwaite. 2002. "Enhancing VHA's Mission to Improve Veteran Health: Synopsis of VHA's Malcolm Baldrige Award Application." *Quality Management in Health Care* 10: 29–37.

Perlin, J., L. E. Kazis, K. M. Skinner, X. S. Ren, A. Lee, W. R. Rogers, A. Spiro III, A. Selim, and D. R. Miller. (2000). Health Status and Outcomes of Veterans: Physical and Mental Component Summary Scores, Veterans SF-36[R], 1999 Large Health Survey of Veteran Enrollees. Executive Report. Department of Veterans Affairs, Veterans Health Administration, Office of Quality and Performance. Washington, DC. May 2000.

Petersen, L. A., S. L. Normand, J. Daley, and B. J. McNeil. 2000. "Outcome of Myocardial Infarction in Veterans Health Administration Patients as Compared with Medicare Patients." *New England Journal of Medicine* 343: 1934–41.

Porell, F. W., and H. B. Miltiades. 2001. "Disability Outcomes of Older Medicare HMO Enrollees and Fee-for-Service Medicare Beneficiaries." *Journal of the American Geriatric Society* 49: 615–31.

Reason, J. 2000. "Human Error: Models and Management." *British Medical Journal* 320: 768–70.

Retchin, S. M., D. G. Clement, L. F. Rossiter, B. Brown, R. Brown, and L. Nelson. 1992. "How the Elderly Fare in HMOs: Outcomes from the Medicare Competition Demonstrations." *Health Services Research.* 27: 651–69.

Riley, G. 2000. "Two-Year Changes in Health and Functional Status among Elderly Medicare Beneficiaries in HMOs and Fee-for-Service." *Health Services Research.* 35: 44–59.

Rosenthal, G. E., P. J. Kaboli, and M. J. Barnett. 2003. "Differences in Length of Stay in Veterans Health Administration and Other United States Hospitals: Is The Gap Closing?" *Medical Care* 41: 882–94.

Schall, L. C., J. M. Buchanich, G. M. Marsh, and G. M. Bittner. 2001. "Utilizing Multiple Vital Status Tracing Services Optimizes Mortality Follow-Up in Large Cohort Studies." *Annals of Epidemiology* 11: 292–6.

Schneeweiss, S., P. S. Wang, J. Avorn, and R. J. Glynn. 2003. "Improved Comorbidity Adjustment for Predicting Mortality in Medicare Populations." *Health Services Research* 38: 1103–20.

Selim, A. J., L. E. Kazis, W. Rogers, S. Qian, J. A. Rothendler, A. Lee, X. S. Ren, S. C. Haffer, R. Mardon, D. Miller, A. Spiro III, B. J. Selim, and B. G. Fincke. 2006. "Risk-Adjusted Mortality as an Indicator of Outcomes: Comparison of the Medicare Advantage Program with the Veterans' Health Administration." *Medical Care* 44: 359–65.

———. 2007. "Change in Health Status and Mortality as Indicators of Outcomes: Comparison between the Medicare Advantage Program and the Veterans Health Administration." *Quality of Life Research* 16: 1179–91.

Shalabh 2001. "Consistent Estimation through Weighted Harmonic Mean of Inconsistent Estimators in Replicated Measurement Error Models." *Economic Reviews* 4: 507–10.

Shaughnessy, P. W., R. E. Schlenker, and D. F. Hittle. 1994. "Home Health Care Outcomes under Capitated and Fee-for-Service Payment." *Health Care Finance Review* 16: 187–222.

Vaughan-Sarrazin, M. S., B. Wakefield, and G. E. Rosenthal. 2007. "Mortality of Department of Veterans Affairs Patients Admitted to Private Sector Hospitals for 5 Common Medical Conditions." *American Journal of Medical Quality* 22: 186–97.

Ware, J. E. Jr., M. Kosinski, and S. Keller. 1994. *SF-36 Physical and Mental Health Summary Scales: A Users's Manual.* Boston: The Health Institute, New England Medical Center.

———. 1995. *SF-12: How to Score the SF-12 Physical and Mental Health Summary Scores.* Boston: The Health Institute, New England Medical Center.

Ware, J. E. Jr, M. S. Bayliss, W. H. Rogers, M. Kosinski, and A. R. Tarlor. 1996. "Differences in 4-Year Health Outcomes for Elderly and Poor, Chronically Ill Patients Treated in HMO and Fee-for-Service Systems. Results from the Medical Outcomes Study." *Journal of the American Medical Association* 276: 1039–47.

Williams, D. R. 1996. "Race/ethnicity and Socioeconomic Status; Measurement and Methodological Issues." *International Journal of Health Services* 26: 483–505.

Wooten, A. F. 2002. "Access to Mental Health Services at Veterans Affairs Community-Based Outpatient Clinics." *Military Medical* 167: 424–6.

Wright, S. M., T. Craig, S. Campbell, J. Schaeter, and C. Humble. 2006. "Patient Satisfaction of Female and Male Users of Veterans Health Administration Services." *Journal of General Internal Medicine* 21 (suppl 3): S26–32.

Young, G. 2006. "Transforming Government: The Revitalization of the Veterans Health Administration" [accessed on November 7, 2008]. Available at http://www.businessofgovernment.org/pdfs/Young_Report.pdf

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix SA1: Author Matrix.

Table SA1: Comparisons of Baseline Patient Characteristics between the Medicare Advantage (MA) Enrollees and Disenrollees* in 2 Years.

Table SA2: Comparisons of Baseline Patient Characteristics between Veterans Who Were Surveyed in the Survey of Healthcare Experiences of Patients (SHEP) and Those Who Were Not Surveyed.

Figure SA1: Baseline PCS and MCS scores for Medicare Advantage (MA) and Veterans Health Administration (VHA) Populations.

Table SA3: Change in Health Status in the Medicare Advantage (MA) and the Veterans Health Administration (VHA).

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.

AIDS Behav (2010) 14:1313–1319
DOI 10.1007/s10461-010-9785-x

BRIEF REPORT

How Researchers Define Vulnerable Populations in HIV/AIDS Clinical Trials

Sohini Sengupta · Bernard Lo · Ronald P. Strauss · Joseph Eron · Allen L. Gifford

Published online: 19 August 2010

Abstract In this study, we interviewed researchers, asking them to define vulnerable populations in HIV/AIDS clinical trials, and provide feedback on the federal regulations for three vulnerable populations. Interview data informed a conceptual framework, and were content analyzed to identify acceptability or disagreement with the regulations. Beginning with several characteristics of vulnerable enrollees identified by researchers, the conceptual framework illustrates possible scenarios of how enrollees could be considered vulnerable in clinical research. Content analysis identified barriers affecting HIV/AIDS researchers' ability to conduct clinical trials with pregnant women, prisoners, and children, for which the regulations specify additional protections. This study challenges current thinking about federal regulations' group-based approach to defining vulnerable populations.

Keywords HIV · AIDS · Vulnerable populations · Ethics · Clinical trials

S. Sengupta (✉)
Center for Faculty Excellence, The University of North Carolina at Chapel Hill (UNC-Chapel Hill), 316 Wilson Library, 3470, Chapel Hill, NC 27599-3470, USA
e-mail: sengups@unc.edu

B. Lo
Department of Medicine, University of California, San Francisco (UCSF), San Francisco, CA, USA

R. P. Strauss
Executive Associate Provost, School of Dentistry and School of Medicine, UNC-Chapel Hill, Chapel Hill, NC, USA

J. Eron
Department of Medicine, Division of Infectious Diseases, Center for AIDS Research (CFAR), UNC-Chapel Hill, Chapel Hill, NC, USA

A. L. Gifford
VA New England Healthcare System, Center for Healthcare Quality, Outcomes and Economic Research, Bedford, MA, USA

A. L. Gifford
Department of Health Policy and Management and Department of Medicine, Boston University, Boston, MA, USA

Introduction

Vulnerability is a central concept in protecting human subjects in research, and the term, *vulnerable populations*, was introduced as part of the guidelines for medical ethics in the 1949 Nuremburg Code, World Medical Association Declaration of Helsinki (most recent update: 2008), and the 1979 Belmont Report to protect human subjects involved in research [1, 2]. The US federal regulations for protection of human subjects, in 45 Code of Federal Regulations Part 46 (45 CFR 46) require special protections for three categories of vulnerable populations—pregnant women, fetuses, and neonates (Subpart B), prisoners (Subpart C), and children (Subpart D) [1]. In addition, 45 CFR 46, in Subpart A (also known as the Common Rule), requires Institutional Review Boards (IRBs) to consider additional protections for those who are "economically," "educationally," or "decisionally" impaired, without specifying these terms [1].

Using a group-based approach—in which individuals are considered vulnerable if they belong to specified groups—has been criticized to be both too narrow and too broad in scope; too narrow because it does not take into account other factors that lead to vulnerability, or persons or populations with multiple vulnerabilities [3, 4], and too broad because some individuals who belong to these categories are not vulnerable in certain types of research [3,

5]. Another criticism with the group-based approach to identifying vulnerable populations is that the federal regulations do not provide adequate guidance about what additional safeguards should be taken with each of the groups identified [3]. For example, protecting vulnerable populations by barring them from participation in certain clinical trials may be doing more harm than good. Indeed, the current conceptualization of vulnerable populations has made access to clinical trials more difficult for underrepresented groups [1].

While the debate concerning defining vulnerable populations continues among bioethics scholars, only a few empirical studies examine how researchers think about and address vulnerability in clinical trials [6–8]. To date, however, no studies have examined issues of vulnerable populations in the context of HIV/AIDS clinical trials, particularly from the perspectives of frontline researchers. HIV/AIDS, unlike other diseases/conditions, includes individuals and populations with a wide range of vulnerability characteristics—including racial/ethnic minorities, women, and/or injecting drug users—not necessarily specified in the current definition and categories [9, 10]. HIV-positive subjects with one or more of these characteristics may not fit in 45 CFR 46's Subparts B,C, and D. Moreover, the terms, "educational, economic, and decisional impairment" in the Common Rule are not specific enough to assist HIV/AIDS clinical trial researchers to identify and protect subjects who they consider vulnerable.

The purpose of this study is twofold: The first part is to take a grounded-theory approach in developing a conceptual framework for understanding which enrollees are considered vulnerable from the perspectives of researchers working in HIV/AIDS clinical trials. The second part of the study explores HIV/AIDS researchers' perspectives on the subparts of 45 CFR 46 to gain a better understanding of the perceived utility and limits of these regulations for protecting pregnant women/fetuses, prisoners, and children from potential risks.

Methods

Sample and Recruitment

A sampling frame of AIDS Clinical Trials was obtained from the AIDS Clinical Trial Information Service (ACTIS), which was a resource of federally and privately-funded AIDS Clinical Trial information through the Division of Acquired Immune Deficiency Syndrome (DAIDS), National Institute for Allergy and Infectious Diseases (NIAID) [11]. A list of Principal Investigators (PIs) was created in an Excel spreadsheet and categorized by Adult or Pediatric AIDS Clinical Trials Groups (ACTGs), and comprised 31 Adult AIDS Clinical Trials Group (AACTG) sites and 18 Pediatric AIDS Clinical Trials Group (PACTG) sites. In addition, 16 HIV Prevention Trials Network (HPTN) PIs were included in the spreadsheet. HPTN—established in 1999 by DAIDS—is an international clinical trials network that develops and tests the safety and efficacy of primarily non-vaccine interventions designed to prevent the transmission of HIV [12].

All AACTG, PACTG, and HPTN PIs were sent an initial letter via regular mail requesting their participation in a telephone interview, and the names and contact information of their co-investigators or study coordinators so that they could be asked to participate. To PIs, co-investigators, and study coordinators who agreed to participate, a consent form was faxed to them; they were asked to sign the consent form and fax it back before the scheduled telephone interview. Several follow-up email or telephone contacts were made to both investigators and study coordinators to maximize the responses; both refusals and non-responses were documented. Once the interview was completed, investigator- and study coordinator-subjects were mailed a $50 incentive, or the $50 was donated to a charity of their choice. This study was approved by the UNC Biomedical Institutional Review Board.

Data Collection

A semi-structured interview guide was developed and used to conduct the audiotaped, telephone interviews with the investigators/study coordinator subjects. The interview guide included two conceptual domains to explore (a) their definitions of what constitutes vulnerability in HIV/AIDS clinical trial populations, and how they see their study population(s) as vulnerable; and (b) what they think of the current categories of vulnerable populations for which there are special protections in the 45 CFR 46.

Data Analysis

All audiotaped interviews from investigator/study coordinator subjects were electronically transcribed into Microsoft Word. Accuracy of the transcription was verified by a member of the research team, and any identifying information in the interviews was redacted to protect the confidentiality of subjects. The transcribed interviews were imported into the qualitative software program, Atlas.ti, v. 5.2. The first phase of qualitative data analysis involved identifying themes from the questions asked, and developing a codebook reflecting a thematic coding structure underlying the conceptual domains. Codes for each theme were assigned to text using Atlas.ti by a pair of coders per transcript, and inter-coder reliability was assessed by having the coders resolve any coding differences between

them. Thus, the first phase of the analytical process yielded discrete and systematically coded textual data.

Development of a Conceptual Framework

To develop the conceptual framework, we extracted coded textual data elicited from the questions (a) "What is your definition of a vulnerable population?", and (b) "In general, how are AIDS clinical trial (or HPTN) study populations considered vulnerable?" These data were reviewed in a 2-step process. The first step identified indicators of vulnerability elicited from these two questions and organized them into three broad categories reflecting social, treatment-related, and research participation-related vulnerabilities. The next step involved examining co-occurrences between the themes from these broad vulnerability categories that informed a conceptual framework to explore how HIV/AIDS clinical trial populations could be vulnerable in *clinical* research.

Investigator/Study Coordinator Views on 45 CFR 46 Subparts B, C, and D

To explore the perspectives of HIV/AIDS clinical trial researchers on the three categories of vulnerable populations for which there are special protections in the federal regulations, coded textual data were extracted from the following set of questions:

"IRBs identify three categories of vulnerable populations for which there are special protections. They are children, pregnant women/neonates/fetuses, and prisoners.

- Do you think each of these groups should be considered vulnerable populations with special protections, why or why not?
- Do these categories work well to help researchers understand vulnerable populations in HIV/AIDS clinical trials? Why or why not?"

Textual data were then collapsed into *why* or *why not* categories for each of the three vulnerable populations to create a matrix.

Results

Sociodemographics

Of the 65 sites (31 AACTG, 18 PACTG, 16 HPTN) contacted from the sampling frame, 27 (42%) sites' PIs agreed to participate: 14 (45%) representing AACTG, 5 (28%) from PACTG, and 8 (50%) from HPTN. Of the 38 sites that did not participate, 36 of the PIs from those sites never responded to our initial or follow-up requests, and two of the PIs declined to participate (no reasons were given). Furthermore, non-respondent PIs disproportionately represented the PACTG. Table 1 presents the sociodemographics of the 38 investigator- and study coordinator-subjects from the 27 participating sites. Investigators primarily were male (67%) and were physicians (72%), while the study coordinators primarily were female (65%) and were nurses (95%). The majority of the investigators and study coordinators were White (87%).

Table 1 Sociodemographics of investigator- and study coordinator-subjects

Variables	Investigators ($N = 18$) *n* (%)	Study coordinators ($N = 20$) *n* (%)
Gender		
Male	12 (67)	7 (35)
Female	6 (33)	13 (65)
Race/Ethnicity		
White, not Hispanic	16 (89)	17 (85)
Black, not Hispanic	–	2 (10)
Asian	2 (11)	1 (5)
Trial type		
AACTG	8 (44)	17 (85)
PACTG	3 (17)	2 (10)
HPTN	7 (39)	1 (5)
Degree(s)		
MD	13 (72)	–
PhD or equivalent	4 (22)	–
Nursing degree (RN, NP)	1 (6)	19 (95)
Master's degree (MA, MS)	–	1 (5)
Location of trials		
US	7 (39)	16 (80)
International	4 (22)	–
Both	7 (39)	4 (20)

Conceptual Framework Defining Vulnerable Populations in HIV/AIDS Clinical Trials

The two initial, open-ended questions asking about how study populations in HIV/AIDS clinical trials are vulnerable elicited an array of indicators that we first organized into three broad vulnerability categories: social (e.g., substance/alcohol abuse, homeless), treatment-related (e.g., few options to treatment, newly diagnosed), and research participation-related (e.g., not understanding consent form, participating for inducement, physician-investigator influence). While organizing these vulnerable indicators into categories was useful, these indicators alone did not fully explain the possible circumstances in which HIV/AIDS clinical trial enrollees could be considered vulnerable in

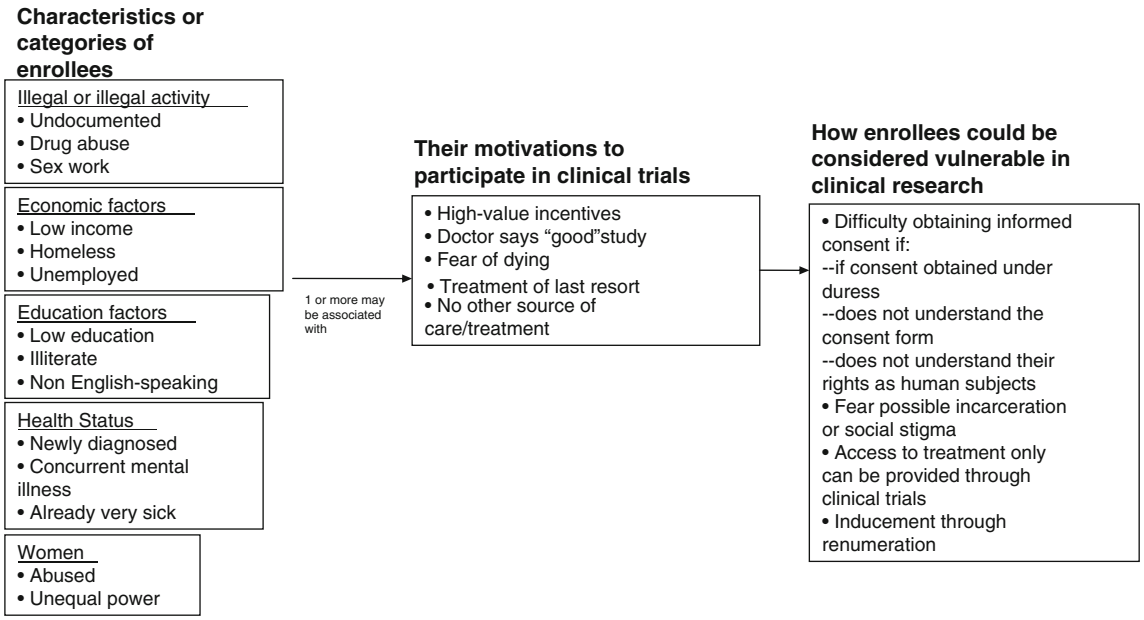


Fig. 1 Conceptual framework for understanding vulnerability in HIV/AIDS clinical trials

clinical research. To answer this question, we explored the thematic co-occurrences among these indicators, which were then used to develop a conceptual framework informing how researchers define vulnerable populations in HIV/AIDS clinical trials (Fig. 1).

The following quotation illustrates an example of the co-occurrences among some of the social, treatment-related, and research participation-related vulnerabilities depicted in Fig. 1; vulnerable population indicators are highlighted in italics. In this example, the study coordinator was sharing a story about a new enrollee who also just found out that he was HIV-positive:

> I'll take an extreme example where I was called into consent a patient…*They needed to be on meds years ago and just found out they're positive*. This individual was tearful through part of our consenting, *just overwhelmed with the new diagnosis, the pills, not understanding that they're not going to be dead in 2 years…And there's a lot of education there but they're only going to take home 10% of what you tell them*…So I mean that is a vulnerable person and I think they're informed as much as we can humanly inform them and they're making a legal consent, but they're vulnerable. *They have their faith in the physician that we're not going to lead them wrong. It's just my doctor said this would be a good study*. (HPTN study coordinator, male)

Relating back to Fig. 1, this quotation reflects the multiple vulnerabilities an enrollee can face in HIV/AIDS in clinical research. This particular enrollee had been sick for while, perhaps was motivated to participate because of the physician-investigator's influence, and believed that he was going to die given the HIV-positive diagnosis. The end result may be that this enrollee agreed to participate, despite understanding very little about the clinical trial during the consent process, because the trial in question may have been the only source of treatment/care available to him at that time.

Investigator and study coordinator subjects also described how their HIV/AIDS clinical trial sites have made efforts to provide special protections to enrollees who have one or more of the vulnerabilities presented in Fig. 1. One special protection that may not be required by IRBs, but was commonly cited by clinical trial researchers, was the implementation of various consent monitoring practices to improve informed consent comprehension at the time of enrollment. In this example, one AACTG female study coordinator stated:

> We do not enroll any study subject that we feel does not have a clear understanding. And we sort of give them like a little quiz…So, for example, we run studies for people that just got diagnosed and just found out they're HIV-positive…If we don't feel like they can handle not only the information of just being diagnosed with HIV and then on top of that a research study, we don't put them in the study.

Other special protections for HIV/AIDS clinical trial enrollees that were described included:

- Informed consent process could be re-visited periodically to make sure enrollees retain comprehension of key information in the clinical trial.

- Having trained staff to provide various services, such as translators to communicate with non-English speakers, social workers to help prevent crises in a patient's life arising from participating in a trial, or guardians present for the cognitively-impaired.
- Having a certificate of confidentiality approved by the IRB, especially for study populations involved in illicit activities.
- Working with high ranking officials and law enforcement officers to ensure research proceeds smoothly when there is illegal activity involved.

Thus, investigator and study coordinator subjects went beyond merely identifying different factors associated with vulnerability to suggesting specific measures to ameliorate the different types of vulnerability their participants faced.

Perspectives on 45 CFR 46 Subparts B, C, and D

Investigator and study coordinator subjects were asked how they felt about the current categories of vulnerable populations for which there are special protections in the Subparts of 45 CFR 46. Table 2 dichotomizes their perspectives by why or why not categories for pregnant women/neonates/fetuses, prisoners, or children should be considered vulnerable.

In general, investigator- and study coordinator-subjects agreed with the need for special protections for the categories of vulnerable populations identified in Subparts B, C, and D of 45 CFR 46 for the following reasons: women/neonates/fetuses because of the potential or known harms to the fetus (25 [66%]); prisoners given their captivity (24 [63%]); and children because of their age (33 [87%]). Yet, the investigators and study coordinator subjects did not view these study populations as always being vulnerable. Their reasons for why they should not be considered vulnerable further demonstrate how they consider vulnerability situational, thus rejecting the regulations' categorizing of these groups as vulnerable in all types of research. The following are sample quotes that illustrate how IRBs sometimes have made it difficult for researchers to enroll potential subjects who are pregnant, prisoners, or children.

For pregnant women:

> Well, the kinds of behavioral research I do, the interventions are applicable or more applicable for pregnant women. There's nothing that we do that would in any way harm the fetus…But I have to explain to the IRB why I'm including them…I don't get it. (HPTN investigator, male)

In this example, the regulations would deem behavioral intervention trials, such as the types described, as having minimal risk to the fetus, but the IRB made it more difficult

Table 2 Investigator and study coordinator subjects' perspectives on 45 CFR 46 subparts B, C, and D

45 CFR 46 subparts	Do you think each of these groups should be considered vulnerable populations with special protections?			
	Why?	*n*	Why not?	*n*
Subpart B: pregnant women/neonates/fetuses	• Experimental drugs could:	30	• Just because a woman is pregnant does not make her vulnerable	1
	1. Cause spontaneous abortion	5	• Becoming pregnant after starting a trial should not be the only reason to rule out a woman's ability to continue in the trial	3
	2. Interfere with fetus development	25		
	• Pregnant women only vulnerable if they are socioeconomically disadvantaged or affected by domestic violence	5	• Pregnant women should not be considered vulnerable in monitoring/observational trials, many prevention trials, and in clinical trials with medications that are proven non-teratogenic	13
	• Hormonal changes affect pregnant women's decision-making	7		
Subpart C: Prisoners	• Prisoners are a captive population. They do not have the ability to make their own choices	24	• Prisoners can make decisions for themselves since they are adults	3
			• Prisoners have better resources in prison than some populations on the outside	2
	• Co-morbid factors may be present, such as high rates of HIV infection or drug use that make them more vulnerable	4	• Entering prison post enrollment, prisoners not vulnerable—they already consented	13
Subpart D: Children	• Children cannot consent for themselves due to their age or lack of maturity	33	• Current age criterion is flawed given that it doesn't take into account cultural and maturity factors when qualifying someone as a child versus an adult	22
	• Parents can consent under duress or desperation and make bad choices for their children	3		

for this investigator's clinical trials site to include pregnant women.

For prisoners:

> We think it's a bad thing [considering prisoners a vulnerable population with special protections]. We would like to be able to continue their involvement in the clinical trial while they were in jail…but the logistics of actually being able to continue to provide research medications while somebody is in jail makes it very difficult. (AATCG investigator, male)

Investigators have to go though additional IRB procedures to be able to conduct clinical trials with prisoners. In this example, the investigator later explains when enrollees should be allowed to continue their participation in an HIV/AIDS clinical trial if they become incarcerated. Indeed, if enrollees are forced to withdraw from the trial, there is no guarantee that the antiretroviral medications they were taking before incarceration would be available to them during their jail time or imprisonment, which could have more harmful implications.

For children:

> In this country maybe if you're 15 and you want to go into a research study your parent or guardian would have to sign the consent as well. I don't think that's realistic in another country. I think that a 15-year-old who has been out working and might be married should be able to sign and consent to a study on their own. (HPTN study coordinator, female)

Similar to the quotation example for pregnant women, the regulations would permit minor assent without parental permission under certain conditions, but this researcher's experience with the IRB may not have allowed it.

Discussion

This exploratory study provides an in-depth and multi-faceted look at the ethical concept of vulnerable populations from the perspectives of researchers involved in HIV/AIDS clinical trials. A conceptual framework was developed that illustrates a combination of vulnerable population characteristics/categories identified from the broad definitions in the Common Rule (e.g., based on socioeconomics), and in prior reviews about HIV-infected and HIV-affected populations [9, 10], but organizes them to show possible circumstances in which HIV/AIDS clinical trial enrollees could be considered vulnerable in clinical research, and for which special protections could be warranted (e.g., consent monitoring in cases where there is educational, cognitive, or treatment-related vulnerabilities). In addition, barriers posed by Subparts B, C, and D of 45 CFR 46 were explored in the context of HIV/AIDS clinical trials, further demonstrating that applying group-based vulnerability can result in unnecessary exclusion of individuals in the vulnerable population categories of pregnant women/fetuses, prisoners, and children. The study's findings are consistent with concerns from bioethicists that vulnerability in research has not been clearly or uniformly defined in the federal regulations on human subjects research, and special protections delineated for select vulnerable populations often have made it difficult for these vulnerable populations to participate in HIV/AIDS clinical trials from which they could benefit [2, 5].

First, the study supports prior recommendations to look at vulnerability as situational, rather then group-based, as is typically used by IRBs in the United States and internationally [3, 4]. At a practical level, we hope that AACTG, PACTG, and HPTN researchers could use the conceptual framework as a guide to identify areas of vulnerability within their respective clinical trial populations, and implement appropriate protections to their informed consent process and/or recruitment procedures. Furthermore, we believe that the factors comprising the conceptual framework are not necessarily distinct to HIV/AIDS and may broaden the conceptual framework's potential to understanding vulnerable populations for other diseases or conditions, or that are distinct to minority groups (e.g., being undocumented). At an institutional level, we hope that local IRBs could use this conceptual framework to expand their thinking about what makes subjects vulnerable in HIV/AIDS clinical trials, and to consider innovative and alternative ways in which researchers could implement additional protections for clinical trial participants with one or more vulnerabilities.

Second, understanding the concerns among HIV researchers on why pregnant women, prisoners, and children should not be considered vulnerable for all types of clinical research strengthens the argument to examine vulnerability as situational. At an institutional level, we hope that these findings could encourage local IRBs to think of these three vulnerable populations more on a case-by-case basis, particularly in situations where clinical trial participants who represent these vulnerable populations stand to benefit from the results of the research.

This study has two main limitations. For the investigator/study coordinator sample, the response rate of 42% at the site level was low even though all 65 sites (31 AACTG, 18 PACTG, 16 HPTN) were asked to participate. Given that we were unsuccessful in interviewing PIs, other investigators, or study coordinators from 38 of the sites, our conceptual framework may be missing other important factors that contribute to vulnerability in clinical research. This is particularly true for PACTG that may have other vulnerable population indicators more relevant when

conducting clinical trials with children (and adolescents). Second, the conceptual framework for vulnerable populations in HIV/AIDS clinical trials is missing the perspectives coming from HIV/AIDS clinical trial enrollees with respect to how they consider themselves vulnerable. Future research may be useful to better explore vulnerability in clinical research from the enrollees' perspective.

In conclusion, the debate over vulnerable populations usually happens among bioethics scholars with little or no input from biomedical researchers about what they think, including if they agree with the current definitions, and their concerns with IRB regulations associated with vulnerable populations. The conceptual framework for understanding vulnerable populations in HIV/AIDS clinical trials was developed in this study from the perspectives of HIV/AIDS researchers to help us move beyond the broad categories of "economically," "educationally," or "decisionally" impaired, or the narrowly focused vulnerable categories of pregnant women, prisoners, and children for which there are special protections in 45 CFR 46. Indeed, the conceptual framework illustrates other plausible relationships of vulnerability, and this study's participants also identified interventions to address certain types of vulnerability, to which IRBs should pay greater attention. Future work on understanding vulnerable populations could expand the conceptual framework to include vulnerable population indicators for other diseases or conditions, and the conceptual framework could be used as a guide to develop interventions that address situational vulnerability in clinical research.

Acknowledgments Research was funded by a Career Developmental Award in Research Ethics through the National Institute of Allergy and Infectious Diseases (NIAID), Grant# K01AI055247-05. The research also was supported by the UNC-Chapel Hill Center for AIDS Research (CFAR), NIH Grant #P30 AI50410.

References

1. Park SS, Grayson MH. Clinical research: protection of the "vulnerable"? J Allergy Clin Immunol. 2008;121(5):1103–7. PMID: 18313131.
2. National Institutes of Health. Office of Human Subjects Research. Regulations and ethical guidelines (June 2005). http://ohsr.od.nih.gov/guidelines/index.html. Accessed 27 July 2010.
3. National Bioethics Advisory Commission (NBAC). Ethical and policy issues in research involving human participants, vol I (Aug 2001). http://bioethics.georgetown.edu/nbac/human/overvol1.pdf. Accessed 27 July 2010.
4. Shivas T. Contextualizing the vulnerability standard. Am J Bioeth. 2004;4(3):84–6. PMID: 16192157.
5. Levine C, Faden R, Grady C, Hammerschmidt D, Eckenwiler L, Sugarman J. Consortium to examine clinical research ethics. The limitations of "vulnerability" as a protection for human research participants. Am J Bioeth. 2004;4(3):44–9. PMID: 16192138.
6. Newton SK, Appiah-Poku J. Opinions of researchers based in the UK on recruiting subjects from developing countries into randomized controlled trials. Dev World Bioeth. 2007;7(3):149–56. PMID: 18021120.
7. Mason VL, Shaw A, Wiles NJ, et al. GPs' experiences of primary care mental health research: a qualitative study of the barriers to recruitment. Fam Pract. 2007;24(5):518–25. PMID: 17698979.
8. Ensign J, Ammerman S. Ethical issues in research with homeless youths. J Adv Nurs. 2008;62(3):365–72. PMID: 18426461.
9. De Santis J. Exploring the concepts of vulnerability and resilience in the context of HIV infection. Res Theory Nurs Pract. 2008; 22(4):273–87. PMID: 19093664.
10. Peragallo N, Gonzalez RM. Nursing research and the prevention of infectious diseases among vulnerable populations. Annu Rev Nurs Res. 2007;25:83–117. PMID: 17958290.
11. Katz DG, Dutcher GA, Toigo TA, Bates R, Temple F, Cadden CG. The AIDS Clinical Trials Information Service (ACTIS): a decade of providing clinical trials information. Public Health Rep. 2002;117(2):123–30. PMID: 12356996.
12. HIV Prevention Trials Network, Research Sites. http://www.hptn.org/hptn_structure/ResearchSites.asp. Accessed 27 July 2010.

A Transitional Opioid Program to Engage Hospitalized Drug Users

Christopher W. Shanahan, MD, MPH[1,2], Donna Beers, RN, BSN, CARN[2,4], Daniel P. Alford, MD, MPH[1,2], Eileen Brigandi[4], and Jeffrey H. Samet, MD, MA, MPH[1,2,3,4]

[1]Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, Boston, MA, USA; [2]Boston Medical Center, Boston, MA, USA; [3]Department of Social and Behavioral Sciences, Boston University School of Public Health, Boston, MA, USA; [4]Division of Substance Abuse Prevention and Treatment Services, Boston Public Health Commission, Boston, MA, USA.

BACKGROUND: Many opioid-dependent patients do not receive care for addiction issues when hospitalized for other medical problems. Based on 3 years of clinical practice, we report the Transitional Opioid Program (TOP) experience using hospitalization as a "reachable moment" to identify and link opioid-dependent persons to addiction treatment from medical care.

METHODS: A program nurse identified, assessed, and enrolled hospitalized, out-of-treatment opioid-dependent drug users based on their receipt of methadone during hospitalization. At discharge, patients transitioned to an outpatient interim opioid agonist program providing 30-day stabilization followed by 60-day taper. The nurse provided case management emphasizing HIV risk reduction, health education, counseling, and medical follow-up. Treatment outcomes included opioid agonist stabilization then taper or transfer to long-term opioid agonist treatment.

RESULTS: From January 2002 to January 2005, 362 unique hospitalized, opioid-dependent drug users were screened; 56% (n = 203) met eligibility criteria and enrolled into the program. Subsequently, 82% (167/203) presented to the program clinic post-hospital discharge; for 59% (119/203) treatment was provided, for 26% (52/203) treatment was not provided, and for 16% (32/203) treatment was not possible (pursuit of TOP objectives precluded by medical problems, psychiatric issues, or incarceration). Program patients adhered to a spectrum of medical recommendations (e.g., obtaining prescription medications, medical follow-up).

CONCLUSIONS: The Transitional Opioid Program (TOP) identified at-risk hospitalized, out-of-treatment opioid-dependent drug users and, by offering a range of treatment intensity options, engaged a majority into addiction treatment. Hospitalization can be a "reachable moment" to engage and link drug users into addiction treatment.

Received March 20, 2009
Revised September 8, 2009
Accepted February 10, 2010
Published online March 17, 2010

KEY WORDS: harm reduction; opioid dependence; methadone; addiction treatment.
J Gen Intern Med 25(8):803–8
DOI: 10.1007/s11606-010-1311-3

INTRODUCTION

Opioid-dependent individuals are frequent users of hospital services for acute medical conditions.[1–3] When hospitalized, they are often not engaged in addiction treatment.[4] They avoid hospitalization for fear of withdrawal,[5] and if hospitalized, resume drug use at discharge.[6] However, hospitalization provides an opportunity to improve and coordinate addiction treatment. A hospital-based program could identify opioid-dependent patients, engage them in addiction treatment, and mitigate high-risk behaviors. Hospitalization is a "reachable moment"—to engage out-of-treatment individuals whose acute illness may render them willing to consider addiction treatment.[7,8]

Methadone is recommended for acute withdrawal in opioid-dependent patients to reduce early hospital departures and facilitate acute treatment.[4,9,10] However, brief methadone exposure does not improve low abstinence rates (e.g., 80% of opioid-dependant individuals relapse within 1 year of detoxification).[11,12] Moreover, opioid agonists administered during hospitalization do not result in adequate ongoing abstinence after discharge.[8]

Research programs targeting hospitalized opioid-dependent patients have combined engagement with intensive, structured substance use treatment.[13–15] However, patients ambivalent about formal addiction treatment may be disinclined to enroll or remain in these types of programs. We created a clinical model to improve comprehensive health and lifestyle outcomes (e.g., linkage and adherence to treatment, reduction in unhealthy substance use behaviors) and promote low-threshold access that might engage reluctant patients. In 2001, the Transitional Opioid Program (TOP) was created for out-of-treatment, opioid-dependent patients hospitalized in an urban teaching hospital in coordination with but independent of an affiliated Opioid Treatment Program (OTP). TOP had three aims: (1) improve access to opioid addiction treatment; (2) provide risk reduction strategies to prevent HIV, hepatitis, and sexually transmitted diseases; (3) increase hospital discharge plan adherence. TOP identified hospitalized out-of-treatment, opioid-dependent patients and linked them to outpatient,

interim opioid agonist addiction treatment, medical care, risk-reduction services, and individualized case management. Components of this clinical program and the initial 3-year experience are described.

METHODS

The program model was based on a conceptual framework that included the following important components: (1) interim opioid replacement therapy; (2) individualized case management; (3) group public health education; (4) principles of motivational interviewing and harm reduction.

Interim Opioid Agonist Treatment

Opioid agonists facilitated patient engagement in the program by mitigating withdrawal, craving, and illicit drug procurement. The program worked to reduce risk of harm from injection drug use and permit participants to reflect on their circumstances, consider behavior change, and formulate goals.

Individualized Case Management

Individualized case management marshaled resources for vulnerable participants temporarily stabilized on opioid-agonist therapy. After hospitalization, patients face additional barriers to recovery (e.g., lack of long-term treatment availability, lack of insurance, homelessness). Case management provided a safe environment for both formal and informal addiction counseling. Unstructured interactions provided opportunities to address ambivalence to counseling.

Group Health Education

Group health education was an essential program component as patients face many risk-laden decisions.[16,17] Drug users transmit information via "word-of-mouth" and may use knowledge gained through health education to reduce risky behaviors.[18] Group education maximized staff efficiency and reinforced therapeutic messages.

Principles of Motivational Interviewing and Harm Reduction

The program used principles recommended by the Institute of Medicine including engaging patients in all states of readiness to change, setting intermediate goals, working collaboratively with patients towards them, and responding to individual needs.[19] This approach was used to educate participants about prevention of sexually transmitted infections, increase linkage to mental health and primary medical care, and enhance adherence to medical treatment. Intermediate behavioral goals and harm reduction methods were employed to enhance readiness to change and decrease medical complications.[20,21]

The program nurse made frequent "check-in" visits during the hospitalization because experience suggested that repeated low-pressure engagement combined with motivational interviewing methods increased enrollment and enhanced outcomes for ambivalent participants. Incremental progress by achieving intermediate outcomes was considered productive and consistent with the transtheoretical model of behavior change.[22] Any reduction of harmful behaviors or increased involvement in treatment or assistance services was positively emphasized.[23]

Transitional Opioid Program Phases

Phase 1. Inpatient—Identification, Screening, Assessment, and Enrollment. Phase 1 (during hospitalization) included identification, screening, comprehensive medical and psychosocial assessment, social service evaluation, daily visitation, substance abuse treatment education, and methadone induction and stabilization. The program nurse employed clinical judgment while determining program eligibility in interested individuals.

Phases 2. Outpatient Days 1 Through 30—Stabilization and Maintenance. Phase 2 (at the OTP during days 1 to 30 after hospital discharge) included methadone dose reassessment and titration, individualized case management, and comprehensive care planning. The program nurse transferred medication and medical information (last methadone dose, admission documentation, laboratory results, PPD status, and urine toxicology) to the OTP. Patients received daily observed methadone at the OTP up to a maximum daily dose (80 mg) to relieve withdrawal symptoms and opioid craving. The dose of 80 mg was the highest dose from which a 60-day detoxification was estimated to be tolerable. The first day of outpatient methadone administration, the program nurse oriented patients to the OTP and daily administration staff.

Participation in case management, education emphasizing risk reduction, health education, and formal addiction counseling was encouraged but not required. Supervised urine drug testing and medical follow-up were recommended. The program nurse monitored patients, offered support and informal treatment counseling, made referrals, and managed psycho-social crises. A physician met participants the first week at the OTP. The program nurse facilitated weekly public-health oriented educational group discussions at the OTP addressing common challenges (e.g., relapse prevention). Participants also met the program nurse weekly for individual 15-min "check-in" sessions. Methadone was administered within 15 min of the group session to encourage attendance. OTP nurses monitored participants daily during methadone administration and were reminded by an electronic alert to direct participants to keep their weekly "check-in" appointment with the program nurse.

Phase 3. Outpatient Days 31 Through 90—Taper or Titration and Transition to Long-Term Addiction Treatment. Phase 3 (days 31-90) included initiation of a 60-day methadone taper or preparation for transfer to another OTP. The program nurse provided both scheduled and drop-in counseling sessions during Phases 2 and 3. As case manager, the program nurse

supported and guided participants, helping them clarify, define, and achieve personal treatment goals. Participants set measurable, achievable goals and identified action steps to meet them (e.g., decrease methadone dose 5 mg per week, keep next doctor appointment). The plan was reviewed weekly and revised based on progress, physical and mental health status, housing and legal status. The program nurse assisted and advocated for the participants to ensure that action steps were met as planned. Participants, in consultation with the program nurse, decided to taper or transfer to a long-term OTP based on individual preference, availability of treatment slots, insurance status, staff recommendations, and employment and family issues.

Outpatient Objectives During Phase 2 and 3. Phases 2 and 3 addressed educational issues, behaviors, and service utilization. Examples included discharge medical treatment adherence (obtaining prescriptions and medical follow-up); harm reduction (needle sharing avoidance, needle exchange program enrollment, vein care, overdose prevention); condom use; HIV counseling and testing; hepatitis C and HIV health education; and addiction treatment education (acupuncture, community resources, methadone, relapse prevention, smoking cessation, recovery tools, and 12-step groups).

Design and Implementation

The Program Nurse. The program nurse consistently engaged patients from hospital enrollment to program completion. Daily visits helped patients establish trust and rapport, address expectations and concerns, and optimize methadone dosing. The program nurse (DB), available to patients and all clinical staff by pager, was supervised by the program physician director (CWS). During Phase 2 and 3, the nurse reviewed treatment plans, assessed dosing, answered questions, and provided emotional support for participants at the OTP. OTP physicians oriented participants to opioid agonist treatment, performed assessments, and collaborated with the program nurse on the treatment plan. The program nurse provided ongoing education to the OTP clinic nursing staff about the philosophy and policies of the program to maintain support for TOP as a distinct program with specific goals within the larger OTP.

To determine program eligibility, hospitalized patients were screened for receipt of daily low-dose methadone (20–40 mg). Patients were identified through the inpatient computerized medication ordering system. The program nurse reviewed individual medical records and discussed the potential for enrollment with clinical staff. Eligibility criteria included: (1) active opioid use and dependence; (2) not currently enrolled in an opioid treatment program; (3) no chronic use of non-prescribed benzodiazepines; (4) no current alcohol-dependence; (5) no active psychosis or suicidal/homicidal ideation. Program participation did not guarantee the opportunity for long-term OTP enrollment. Methadone dosing in the OTP program was based on federal regulations authorizing up to 120 days of methadone without concurrent routine drug testing, counseling or rehabilitation services for individuals awaiting comprehensive treatment.[24]

Outcome Definitions. Important medical and psychosocial patient outcomes were determined after program design and implementation and were reported for descriptive purposes only. A comparison group was not available. Outcomes were defined in terms of whether or not treatment was provided to participants. All outcomes were classified as "Treatment Provided" (e.g., enrolled in long-term OTP), "Treatment Not Provided" (e.g., loss to follow-up), or "Treatment not Possible" because medical or psychiatric issues took precedence (e.g., too medically or mentally ill or incarcerated).

Data Collection. Data from the program case-management database and the hospital's information system were extracted and transferred to a research database. Unique patients from the program's first 3 years are reported. The Institutional Review Board of Boston Medical Center approved this study.

RESULTS

Patient Characteristics

Between January 2002—January 2005, 362 unique patients were screened from admissions to the medical service of Boston Medical Center that received methadone for opiate withdrawal treatment and were neither enrolled in an OTP nor prescribed methadone for pain control (Fig. 1). Average hospitalization was 5.7 days (SD 7.3; range 1-76 days). Average daily in-patient census (Phase 1) was 2, and average program census (Phase 1-3) was 17. The average length of participation

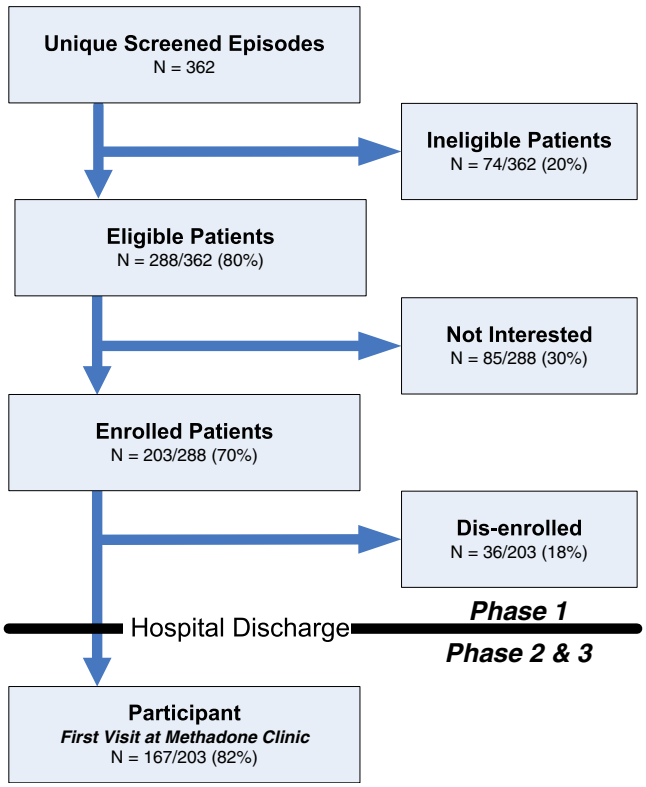


Figure 1. Screening and enrollment schema of the transitional opioid program.

in the entire program (Phases 1-3) was 60 days (SD 38; range, 1-154 days).

Unique Screened Patients

Of the 362 unique screened patients, there were 67% males: 50% White, 45% African-American, or 5% other; 24% self-reported Hispanic ethnicity. Mean age was 40 years. Medical conditions typical for this population were noted.[1,2] Housing information available from a subset of patients assessed in the latter period of follow-up revealed 60% (69/115) "any homelessness in the previous 6 months," 44% (51/115) "currently homeless," and 65% (75/115) "anticipated homelessness in the next 6 months."

Of the screened patients, 20% (74/362) were ineligible for the following reasons: active benzodiazepine abuse 24% (18/74); alcohol dependence with active alcohol use 24% (18/74); unstable psychiatric co-morbidity 14% (10/74); opioid use for less than 1 year 12% (9/74); medical illness severity 9% (7/74); non-daily opioid use 7% (5/74); other reasons 9% (7/74).

Eligible Patients

Of the 362 screened patients, 80% (288) were eligible for enrollment. However, 30% (85/288) declined program enrollment because they "desire residential treatment" 48% (41/85); "oppose methadone treatment" 24% (20/85); "have no interest in treatment at this time" 15% (13/85); "live too far away from the methadone clinic" 9% (8/85); or "want AA/NA only" 4% (3/85). Overall, 89% of those eligible (255/288) reported interest in addiction treatment.

Patients Enrolled in Hospital (Phase 1)

Of the remaining 203 patients that were eligible and accepted enrollment, 82% (167/203) became participants at the OTP clinic after hospitalization. Among the 18% (36/203) who "dropped-out" (failed to transition to the OTP—Phase 2), 44% (16/36) did not appear for the first dose, 22% (8/36) left the hospital "against medical advice," 11% (4/36) became too ill (e.g., transferred to ICU or long-term nursing facility), 8% (3/36) were discharged to another facility, and 14% (5/36) dropped out for other reasons.

Overall Outcomes of Participants Enrolled in Hospital (Phase 1)

Of 203 participants initially enrolled during hospitalization, treatment was provided to 59% (119/203), treatment was not provided to 26% (52/203), or treatment was not possible 16% (32/203).

Short-Term Substance Abuse Treatment Outcomes of Program Participants (Phase 2 and 3)

Among the 203 enrolled participants who entered Phase 2 and had treatment provided, 35% (71/203) enrolled in a long-term OTP, 15% (31/203) completed methadone taper, 4% (9/203) entered outpatient or residential substance abuse treatment, and 2% (5/203) entered an inpatient detoxification facility.

Among 52 participants initially enrolled but who were not provided treatment, 46% (24/52) did not show at the methadone clinic or left the hospital against medical advice, and 54% (28/52) did not show at the OTP clinic for 14 consecutive days or were discharged for behavioral issues.

Other Short-Term Outcomes

Phase 2 participants (n = 167) attained other outcomes including: obtained discharge prescriptions, 56% (94/167); attended primary care appointment, 54% (90/167); attended 2 or more group counseling sessions, 50% (84/167); enrolled in a needle-exchange program, 17% (28/167); attended a 12-step program, 16% (27/167); became employed, 16% (27/167).

DISCUSSION

The Transitional Opioid Program identified, recruited, engaged, and linked hospitalized out-of-treatment opioid-dependent patients to addiction care using interim opioid agonist treatment, individualized case management, and both scheduled and drop-in counseling. Most enrolled participants, 82% (167/203), presented for treatment at the OTP clinic after hospital discharge suggesting that an unmet need was addressed. The program employed minimal enrollment standards that did not require commitment to long-term OTP.

Other investigators have studied methods to engage opioid-dependent patients in addiction treatment. A randomized clinical trial targeting out-of-treatment opioid users identified and linked 126 selected individuals to methadone maintenance using four strategies: case management, free treatment voucher, case management plus voucher, or usual care for 6 months. At 3 months, long-term OTP enrollment was "usual care" (11%), case management alone (47%), treatment voucher (89%), and both case management and voucher (93%). Enhanced treatment access using a voucher was twice as effective as case management alone in linking hospitalized addicts to drug abuse treatment.[25] Vouchers also enhanced 6-month methadone treatment enrollment rates for patients seeking addiction treatment.[26] Schwartz et al. demonstrated that more than three-quarters of 319 heroin-dependent adults on an OTP waitlist randomly assigned to interim methadone remained engaged and entered a long-term OTP.[27,28]

TOP differs from previous reports in a few important ways. Populations identified by ours and other programs were similar in the range of need and readiness to change; however, TOP engaged non-treatment seeking, opioid-dependent patients ambivalent about substance abuse treatment. Moreover, the program approached ambivalence as a dynamic state and by use of interim opioid agonist therapy provided patients an opportunity to address indecision about addiction treatment and modify their readiness to change. Other "programs" described in the literature consisted of clinical trials that created linked but distinctly separate inpatient and outpatient services that were designed to test different methods and settings for engaging patients. These programs had different

objectives, enforced stricter eligibility requirements, or used longer duration and standard OTP structure. In contrast to reports in the literature, our program encouraged but did not require monthly urine drug testing or formal counseling (i.e., no minimum engagement requirements). Similarly, urine testing, if obtained and positive, did not impact program participation. Moreover, duration of participation in the program was roughly half that of other interim opioid-agonist therapy programs.[7–27] Similar to programs described by Aszalos and O'Toole, TOP participants were hospitalized; however, our program employed a single nurse who performed inpatient initial contact and screening through to final outpatient taper or referral to maintenance treatment in an OTP.[13–15]

The program provided the hospital-based clinical team an option for addressing opioid dependence. The TOP model appears sustainable and replicable using a modest staffing model (one nurse) and basic coordination within existing treatment services. The first 3 years of the program were supported by funding from the Massachusetts Department of Public Health. In year 4, the Boston Public Health Commission assumed program costs. Although data to examine health services implications of the program (e.g., hospital re-admissions) were not available, such analyses warrant further study.

Limitations

The program was not available for recruitment at all times but rather only 4 days/week, 8 h/day nor on all services (i.e., only medical service). The TOP model represents a single health care system's experience and awaits replication. Long-term outcomes of the model could not be determined because subjects did not receive long-term follow-up. Use of electronic medication reports was important and enhanced efficiency, but was not essential to the TOP model, and other enrollment screening methods could be devised to facilitate replication. Quantification of the relative improvement in addiction treatment and other outcomes was not possible because a comparison group was not available. However, most participants were not seeking treatment when hospitalized, and few would have received addiction treatment had they not been enrolled in the program.

CONCLUSIONS

The Transitional Opioid Program (TOP) program model is based on collaboration between a traditional acute inpatient facility and an outpatient addiction treatment program. This model of a transitional opioid program (TOP) engaged and linked many opioid-dependent patients to appropriate addiction and medical care by providing interim opioid agonist treatment while offering a range of treatment intensity options along with exposure to case management and health education focused on personal risks. Given the descriptive nature of this report, its findings should be considered preliminary and hypothesis-generating, and further study will be required to evaluate model efficacy and long-term outcomes.

Acknowledgements: *Preliminary study results were presented by CW Shanahan at the American Association for the Treatment of Opioid Dependence Conference, Washington DC (October 2003), and by JH Samet, in a symposium at the 2007 Addiction Health Services Research conference, Athens, Georgia (October 2007). We appreciate the support of the Massachusetts Department of Public Health—HIV/AIDS Bureau and the Bureau of Substance Abuse Services for funding this pilot project, with particular appreciation to Jean F. McGuire, PhD, former Director of the HIV/AIDS Bureau. We also appreciate the leadership of Mr. Tom Scott, former Director of the Boston Public Health Commission, Division of Substance Abuse Prevention and Treatment Services, and Barbara Dworetzky, MD, for her reviews of the manuscript.*

Conflict of Interest: *None disclosed.*

Corresponding Author: *Christopher W. Shanahan, MD, MPH; Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, Department of Medicine, Boston Medical Center, 801 Massachusetts Avenue, 2nd Floor, Boston, MA 02118, USA (e-mail: cshanaha@bu.edu).*

REFERENCES

1. **Samet JH, Shevitz A, Fowle J, Singer DE.** Hospitalization decision in febrile intravenous drug users. Am J Med. 1990;89(1):53–7.
2. **Stein MD.** Medical consequences of substance abuse. Psychiatr Clin North Am. 1999;22(2):351–70.
3. **Burnam MA, Bing EG, Morton SC, et al.** Use of mental health and substance abuse treatment services among adults with HIV in the United States. Arch Gen Psychiatry. 2001;58(8):729–36.
4. **Hopper JA, Shafi T.** Management of the hospitalized injection drug user. Infect Dis Clin North Am. 2002;16(3):571–87.
5. **McCoy CB, Metsch LR, Chitwood DD, Miles C.** Drug use and barriers to use of health care services. Subst Use Misuse. 2001;36(6–7):789–806.
6. **Chutuape MA, Jasinski DR, Fingerhood MI, Stitzer ML.** One-, 3-, and 6-month outcomes after brief inpatient opioid detoxification. Am J Drug Alcohol Abuse. 2001;27(1):19–44.
7. **Samet JH, Rollnick S, Barnes H.** Beyond CAGE. A brief clinical approach after detection of substance abuse. Arch Intern Med. 1996;156(20):2287–93.
8. **O'Toole TP, Pollini RA, Ford D, Bigelow G.** Physical health as a motivator for substance abuse treatment among medically ill adults: is it enough to keep them in treatment? J Subst Abuse Treat. 2006;31(2):143–50.
9. **O'Connor PG, Samet JH, Stein MD.** Management of hospitalized intravenous drug users: role of the internist. Am J Med. 1994;96(6):551–8.
10. **Chan AC, Palepu A, Guh DP, et al.** HIV-positive injection drug users who leave the hospital against medical advice: the mitigating role of methadone and social support. J Acquir Immune Defic Syndr. 2004;35(1):56–9.
11. **Simpson DD, Savage LJ.** Treatment re-entry and outcomes of opioid addicts during a 4-year follow-up after drug abuse treatment in the United States. Bull Narc. 1980;32(4):1–10.
12. **Simpson DD.** Treatment for drug abuse. Follow-up outcomes and length of time spent. Arch Gen Psychiatry. 1981;38(8):875–80.
13. **Aszalos R, McDuff D, Weintraub E, Montoya I, Schwartz R.** Engaging hospitalized heroin-dependent patients into substance abuse treatment. J Subst Abuse Treat. 1999;17(1–2):149–158.
14. **O'Toole TP, Pollini RA, Ford DE, Bigelow G.** The effect of integrated medical-substance abuse treatment during an acute illness on subsequent health services utilization. Med Care. 2007;45(11):1110–5.
15. **O'Toole TP, Conde-Martel A, Young JH, Price J, Bigelow G, Ford DE.** Managing acutely ill substance-abusing patients in an integrated day hospital outpatient program: medical therapies, complications, and overall treatment outcomes. J Gen Intern Med. 2006;21(6):570–6.
16. **Sears C, Guydish JR, Weltzien EK, Lum PJ.** Investigation of a secondary syringe exchange program for homeless young adult injection drug users in San Francisco, California, U.S.A. J Acquir Immune Defic Syndr. 2001;27(2):193–201.
17. **Bortolotti F, Stivanello A, Dall'Armi A, Rinaldi R, La Grasta F.** AIDS information campaign has significantly reduced risk factors for HIV infection in Italian drug abusers. J Acquir Immune Defic Syndr. 1988;1(4):412–3.
18. **Friedman SR, Des Jarlais DC, Sotheran JL.** AIDS health education for intravenous drug users. Health Educ Q. 1986;13(4):383–93.
19. Institute of Medicine. A frame work for improving quality. In: Improving the Quality of Health Care for Mental and Substance-use Conditions, ed.

Quality Chasm Series. Washington: National Academy Press; 2006: 56–76.
20. **Miller WR, Rollnick S.** Phase 2: Strengthening commitment to change. In: Motivational Interviewing, ed. Preparing People for Change. New York: Guilford Press; 2002:428.
21. **Marlett GA.** Harm reduction around the world: a brief history. In: Marlett GA, ed. Harm Reduction: Pragmatic Strategies for Managing High-Risk Behaviors. New York: Guilford Press; 1998:390.
22. **Prochaska JO, DiClemente CC, Norcross JC.** In search of how people change. Applications to addictive behaviors. Am Psychol. 1992;47 (9):1102–14.
23. **Marlatt GA.** Basic principles and strategies of harm reduction. In: Marlatt GA, ed. Harm Reduction: Pragmatic Strategies for Managing High-Risk Behaviors. New York: The Guilford Press; 1998:50–52.
24. Federal Register, title 21, 1993. Codified at 58 CFR 496, pt 291.
25. **Sorensen JL, Masson CL, Delucchi K, et al.** Randomized trial of drug abuse treatment-linkage strategies. J Consult Clin Psychol. 2005;73 (6):1026–35.
26. **Barnett PG, Masson CL, Sorensen JL, Wong W, Hall S.** Linking opioid-dependent hospital patients to drug treatment: health care use and costs 6 months after randomization. Addiction. 2006;101(12):1797–1804.
27. **Schwartz RP, Highfield DA, Jaffe JH, et al.** A randomized controlled trial of interim methadone maintenance. Arch Gen Psychiatry. 2006;63 (1):102–9.
28. **Schwartz RP, Jaffe JH, Highfield DA, Callaman JM, O'Grady KE.** A randomized controlled trial of interim methadone maintenance: 10-month follow-up. Drug Alcohol Depend. 2007;86(1):30–6.

Health Serv Outcomes Res Method (2009) 9:234–251
DOI 10.1007/s10742-010-0056-0

Benefit-of-the-doubt approaches for calculating a composite measure of quality

Michael Shwartz · James F. Burgess · Dan Berlowitz

Received: 6 August 2009 / Revised: 24 February 2010 / Accepted: 26 February 2010 /
Published online: 20 March 2010

Abstract Standard approaches for determining weights when calculating a composite measure of health care quality from individual quality indicators (QIs) include equal weighting, opportunity-based weights, and judgment-based weights. Benefit-of-the-doubt approaches have not been used in the health services area, though one has been used to calculate composite measures for profiling countries. Underlying these approaches is the assumption that relative performance on a set of indicators is, at least to some extent, a revealed preference by the organizational unit about the relative importance of the indicators. A benefit-of-the-doubt approach recognizes these revealed preferences by assigning higher weights to indicators on which performance is better and lower weights to indicators on which performance is poorer. We consider two benefit-of-the-doubt approaches. The first uses simple linear programming (LP) models; the second uses data envelopment analysis (DEA), the way in which the benefit-of-the-doubt approach has been previously implemented. In both cases, constraints are added to limit weight adjustments to some percentage of policy-determined baseline weights. Using both standard and benefit-of-the-doubt approaches, composite scores are calculated from data on five QIs from 32 Department of Veterans Affairs (VA) nursing homes. We examine the tradeoff between the

M. Shwartz · J. F. Burgess
Center for Organization, Leadership and Management Research, VA Boston Healthcare System (152M), 150 South Huntington Avenue, Boston, MA 02130, USA

J. F. Burgess
e-mail: James.Burgess@va.gov

M. Shwartz (✉)
School of Management, Boston University, 595 Commonwealth Avenue, Boston, MA 02215, USA
e-mail: mshwartz@bu.edu

D. Berlowitz
Center for Health Quality, Outcomes and Economic Research, Bedford VA Hospital, 200 Springs Road, Bedford, MA 01730, USA
e-mail: dberlow@bu.edu

J. F. Burgess · D. Berlowitz
School of Public Health, Boston University, 715 Albany Street, T3W, Boston, MA 02118, USA

level of allowable weight adjustment and impact on facility rankings. If weights are constrained to be within 75% of baseline weights, all approaches identify pretty much the same high performing facilities. Weights from benefit-of-the-doubt approaches, because they are able to reflect local preferences and conditions, should be attractive to facilities and, in a collaborative environment, to policy makers.

Keywords Composite measures · Benefit-of-the-doubt approaches · Health care quality · Performance measurement · Linear programming · Data envelopment analysis

1 Introduction

As efforts to improve measurement of quality of care expand, an increasing number of quality indicators (QIs) are being used to monitor progress. These QIs often capture different dimensions of quality that reflect the multiple objectives of provider organizations. However, when assessing overall performance, it is useful to aggregate individual QIs into a composite score (Institute of Medicine 2006). A composite score provides a useful summary of the extent to which top management has created a culture of quality and designed processes to ensure quality throughout the organization. Payers and consumers can use composite scores to compare performance across providers, and policy makers can use them to design incentives to encourage high quality, cost-efficient care.

Many different types of approaches for determining composite scores have been studied in the health services literature (Landrum et al. 2000; Lied et al. 2002; Zaslavsky et al. 2002; Jacobs et al. 2005; Jha et al. 2005; Werner and Bradlow 2006; Agency for Healthcare Research and Quality 2006; O'Brien et al. 2007; Lindenauer et al. 2007; Caldis 2007; Reeves et al. 2007; Shwartz et al. 2008). These approaches can be broadly distinguished by the way in which a construct (i.e., composite score) relates to the individual indicators which comprise it. There are two different types of relationships: first, the construct can be viewed as causing the indicators; or, second, the construct can be viewed as being formed from the indicators (Edwards and Bagozzi 2000). The first type of relationship is called *reflective* to indicate that individual indicators reflect or are manifestations of the underlying construct. A classic example of a reflective construct comes from educational testing: a student's underlying ability is reflected in answers to examination questions. The second type of relationship is called *formative* to indicate that the construct is formed from, or defined by, the individual indicators. Feinstein (1999) has highlighted this distinction in clinical medicine. Clinicians are often interested in combining multiple uncorrelated dimensions into a single score. The Apgar score, which measures the health of newborns, is an example of a widely used formative score.

Deciding whether a construct should be viewed as reflective or formative is based in part on conceptual considerations and in part on empirical evidence (Jarvis et al. 2003). If QIs are correlated, quality can be viewed as a reflective construct. In this case, factor analysis and other types of latent variable models are often used to estimate composite scores. Often, however, QIs are not highly correlated (Gandhi et al. 2002; Jha et al. 2005; Berlowitz et al. 2005; Ryan et al. 2009). In fact, the general trend in health services is to add QIs that broaden the definition of quality and reflect its different dimensions, not to add measures that are highly correlated with existing measures. Even though measures are not highly correlated, the QIs can be combined to create a composite score (a formative construct) by multiplying individual QIs by weights and then summing the results. In this paper, we treat the composite score as a formative construct.

Conceptually, one would like to weight QIs in a formative scale based on their contribution to overall patient health or well-being. However, there is rarely an empirical basis for determining the effect of changes in QIs on these ultimate patient outcomes. As a result, a number of practical approaches for weighting have been proposed, including equal weights, weights based on expert judgment, and weights based on the relative variance of each QI. When QIs are proportions (e.g., the proportion of eligible people who receive some intervention or who experience some adverse event), a common approach is denominator-based or opportunity-based weighting: each QI is weighted by the ratio of number eligible for the indicator to the sum of the numbers eligible for all indicators. None of these approaches is based on a set of underlying priorities or principles related to equity or efficiency and, if pressed, can be hard to defend.

Nardo et al. (2005) in their review of methods for constructing composite indicators discuss the "benefit-of-the-doubt" approach. As described by Cherchye et al. (2007), the conceptual basis for this approach is the following: relative performance on a set of indicators is a revealed preference by the organizational unit about the relative importance of the indicators. A benefit-of-the-doubt approach recognizes these revealed preferences by assigning higher weights to those indicators on which the organizational unit performs well and lower weights to indicators on which it performs less well. Specifically, weights are assigned in a way to optimize the composite measure, subject to a set of specified constraints. As noted by Cherchye et al. (2007, p 5), benefit-of-the-doubt weights "can be connected to a game-theoretic set-up: they can be conceived of as Nash equilibria in an evaluation game between a regulator and an organization." (Semple 1996). The benefit-of-the-doubt approach has been used to assess the relative performance of countries, mainly by researchers in Europe. For example, it has been used to assess macroeconomic performance (Lovell 1995), to reweight components of the Human Development Index (Mahlberg and Obersteiner 2001; Despotis 2005), to assess sustainable development (Cherchye and Kuosmanen 2004), and to evaluate technology achievement (Cherchye et al. 2008).

In the above examples, the benefit-of-the-doubt approach has been implemented through the use of Data Envelopment Analysis (DEA) (Charnes et al. 1978). DEA has been widely used to measure the performance efficiency of organization units (Cooper et al. 2007), where efficiency is the ratio of the weighted sum of different outputs to the weighted sum of inputs. For each organizational unit, weights for each of the outputs and inputs are chosen to optimize efficiency subject to the constraint that the efficiency measure is no greater than 1 for any of the units under consideration. The resulting efficiency measure for a specific organizational unit is relative to the best performing organizational unit(s), that is, the unit or units that are able to achieve an efficiency of 1 using the set of weights that optimize the specific unit's performance. Charnes et al. (1978) showed how this model can be converted to a linear programming model, which can be easily solved.

Benefit-of-the-doubt approaches have not been used to create composite measures in the health care area. However, in a system in which policy makers desire collaborative relationship with providers, the benefit-of-the-doubt approach has some appeal. Essentially, policy makers are saying to providers "We recognize that your relative performance on a set of QIs may reflect managerial decisions based upon the nature of your patients, staff, facility or local environment. Therefore, we will adjust to some extent our set of policy-determined baseline weights and assign somewhat higher weights to those QIs on which you do relatively well and somewhat lower weights on those QIs on which you do relatively less well." Since, within constraints resulting from the amount of adjustment

allowed, weights are assigned to optimize each facility's performance, facilities have less basis for complaints than if the initial policy-determined weights were used.

A criticism sometimes raised concerning benefit-of-the-doubt approaches is that once each organizational unit uses different weights, comparisons across units are no long possible. In response, consider students taking a course in which there are 5 different assignments, each of which is of equal importance. A common grading option is to allow students to drop the assignment with the lowest grade before calculating the final score. In this case, the weighting scheme is zero for the dropped assignment and one-fourth for the other assignments. A variation of this approach, which reduces the chance students will ignore 20% of the course material, is to allow students to place less weight on one of the assignments, e.g., only 12% weight on the assignment with the lowest score and 22% weight on the other 4 assignments. Assuming the assignment dropped (or down-weighted) occurs at random across students, only 20% of the class will be using the same set of weights. Nevertheless, there is no hesitancy in calculating a final score based on the 4 (or 5 if down-weighting is used) assignments and then assigning a grade based on this score. Hence, the principle of comparison when using a common rule that can result in "self-serving" weights is fairly well-accepted. DEA generalizes this principle. As illustrated by the extensive DEA literature, as long as all units "play under the same set of rules," the fact that the rules result in different weights in no way reduces the ability to fairly compare performance across the units.

There seems little doubt that students like the option of dropping their lowest assignment before calculating their final score. All students' final scores will be higher than if the final score were based on all five assignments. However, since final grading is usually done on a relative basis, it is not clear until all assignments are completed who has benefited and who has not from the policy. Nor is the extent to which the policy changes the relative performance of most students clear. If there is relatively little change in the final grade of most students, the policy of dropping the lowest assignment allows the teacher to "buy" student good will without much impact on final student rankings. If policy makers or regulators are seeking buy-in from providers, similar type policies, which can be implemented through benefit-of-the-doubt approaches, may be a way of "buying" provider good will.

In this paper, we apply benefit-of-the doubt approaches to calculating a composite measure of quality from individual QIs. To set the context, a particular reporting period has ended and performance on a set of QIs is available. Policy makers want to calculate a composite measure of performance for three possible purposes: (1) to facilitate choice among consumers by providing some "gestalt" sense of how good a facility is; (2) as part of a pay-for-performance program, to reward facilities whose performance has been particularly strong; and (3) to facilitate learning about policies and practices at high performing facilities in order to spread these to other facilities, thus raising the overall level of performance of all facilities. The weights that are used to calculate the composite measure for the recently completed period are of relevance only in that period. They will not be used prospectively to determine overall performance in the next reporting period.

We begin by describing several standard approaches currently used by central policy makers to establish weights when calculating a composite measure from QIs that are proportions. We then consider two benefit-of-the-doubt approaches. As noted, the benefit-of-the-doubt approach has been implemented previously using DEA. DEA does not require that specific outputs or inputs be on the same scale (e.g., quality and cost can both be outputs), thus eliminating the need to normalize measures before combining them into a

composite. Since different approaches for normalization can result in different rankings (Nardo et al. 2005), avoiding the need for normalization is an advantage of DEA. However, when all outputs are on the same scale, as they are in our example, there is no need to normalize measures. In this case, a reasonable alternative to DEA is a simple linear programming (LP) model that optimizes performance subject to the constraint that the sum of the weights is one. As we illustrate, the simple LP model is much more transparent than DEA. In what follows, we consider implementing the benefit-of-the-doubt approach using both simple LP and DEA. For both approaches for most of the analyses, we require that the weights remain within some range of baseline weights, thus limiting the extent of adjustment. We describe the models used to find the optimal weights for both approaches. Then, using data on 5 QIs from a sample of 32 nursing homes, we calculate the weights and composite scores using the different approaches. Finally, we examine the extent to which the relative performance of the 32 facilities is affected by the level of allowable weight adjustments.

2 Methods

2.1 Data

The data for this study are from an original set of 35 Department of Veterans Affairs (VA) nursing homes that were selected to represent a balanced sample of different sizes, locations and quality of care (Berlowitz et al. 2003). Data were available on 5 QIs that reflect changes in patients' status over time. All of these have been used previously as measures of nursing home quality: pressure ulcer development (Berlowitz et al. 1996; Porell et al. 1998; Mukamel 1997; Zimmerman et al. 1995); functional decline (Porell et al. 1998; Mukamel 1997; Zimmerman et al. 1995); behavioral decline (Arling et al. 1997; Porell et al. 1998); mortality (Braun 1991; Porell et al. 1998); and preventable hospitalizations (Carter 2003). Data used in calculating QIs were from semi-annual patient assessments performed for case-mix-based reimbursements. Pressure ulcer development was recorded if a patient who was ulcer free at one assessment had a stage 2 or deeper pressure ulcer at the subsequent assessment. Functional decline was measured by a change between assessments in a score measuring limitations in eating, toileting, and transferring. Behavioral decline was measured by a change in a score measuring extent of verbal disruption, physical aggression, and socially inappropriate behavior. Mortality was recorded if there was a death within 6 months of an assessment regardless of location. Preventive hospitalizations occurred if the patient was admitted to an acute medical unit within 6 months of an assessment for one of 13 conditions identified as a potentially preventable hospitalization (Agency for Healthcare Research and Quality 2004). Risk-adjustment models have been developed for these QIs: pressure ulcer development (Berlowitz et al. 1996), functional decline (Rosen et al. 2001), behavioral decline and mortality (Berlowitz et al. 2005), and preventable hospitalizations (Pizer et al. 2003). For each patient, the models give a predicted probability of the adverse event in a 6-month period based on risk factors at the time of initial assessment.

Data were collected in 1998 from the 35 nursing homes. However, data on all 5 measures were available for only 32 of 35 nursing homes, largely because 3 facilities switched to a new patient assessment instrument during the data collection period. These are the nursing homes considered in this study. For each patient, we know whether each

adverse event occurred (indicated by a 0 or 1) and the predicted probability of the adverse event from the relevant risk-adjustment model. Actual outcomes and predicted probabilities are summed for each facility, resulting in an *observed* number of cases (O) and an *expected* number of cases (E) for each QI. There are different ways in which observed and expected cases might be converted to an outcome measure, all of which have both strengths and weaknesses (Ash, Shwartz and Peköz 2003). We consider the ratio of observed cases to expected, i.e., the *O/E* ratio, a widely-used way of evaluating performance.

2.2 Calculating a composite score

2.2.1 Standard approaches: equal weights and denominator-based weights

Giving each indicator the same weight (an approach we call, not surprising, *equal weights*) has a great deal of intuitive appeal, but has the disadvantage that low prevalence indicators receive the same weight as high prevalence indicators. Denominator-based weights, the approach used by the Centers for Medicare and Medicaid Services (CMS) for combining individual process measures within clinical conditions (Premier 2003), weight indicators based on their relative prevalence. We modify the CMS approach slightly for our purposes. For each QI i at facility f, we know O_{if}/E_{if}. Similar to the CMS approach, we use the denominators as the basis for determining relative prevalence. Specifically, the weight for QI i at facility f is the ratio of the expected number for this QI to the sum of the expected numbers for all QIs, i.e., $E_{if}/\Sigma_i\ E_{if}$. Thus, each QI is weighted by its relative expected prevalence at the facility. More commonly expected QIs at a facility will have higher $E_{if}/\Sigma_i\ E_{if}$ ratios and hence will receive more weight. We refer to these weights as *facility-specific prevalence-based weights*. Note, that the composite score calculated using facility-specific prevalence-based weights $= \sum_i\ (E_{if}/\Sigma_i\ E_{if})^*\ (O_{if}/E_{if}) = \Sigma_i\ O_{if}/\Sigma_i\ E_{if}$. Hence, a reduction of one adverse event, regardless of QI, will have the same effect on the composite score. In this sense, facility-specific prevalence-based weights may be thought of as a form of equal weighting.

A modification of the above approach is to base weights not on facility-specific expected values, but on expected values summed across all facilities. This is the approach the Agency for Healthcare Research and Quality (AHRQ) uses to create composite measures from the individual Patient Safety Indicators. For each facility, the weight assigned to QI i is $\Sigma_f\ E_{if}/\Sigma_{if}\ E_{if}$. We refer to this approach as *overall prevalence-based weights*. For facilities with a patient mix (in terms of eligibility for the QIs) similar to the average, facility-specific and overall prevalence-based weights will be similar; for facilities with increasingly extreme patient mixes, weights will be increasingly different.

2.2.2 First benefit-of-the-doubt approach: simple LP models

The first benefit-of-the-doubt approach constrains the sum of the weights to equal 1. Thus, each weight represents the proportional importance of the QI.

Let w_{if} be the weight assigned to QI i at facility f. Then, the composite score at facility f is $\Sigma_i w_{if}\ (O_{if}/E_{if})$. Since low values of O_{if}/E_{if} represent high performance, facility f wants weights that minimize $\Sigma_i w_{if}\ (O_{if}/E_{if})$. The following linear programming model summarizes facility f's decision problem: find w_{if} to

$$\begin{aligned} &\text{minimize } \Sigma_i w_{if} \left(O_{if}/E_{if}\right) \\ &\text{subject to } \Sigma_i w_{if} = 1 \\ &\qquad w_{if} \geq 0 \text{ for all } i,\ i = 1, \ldots, 5 \end{aligned}$$

In addition, we want to ensure that at least some minimum weight is placed on each QI and that no more than some maximum weight is placed on any QI. We analyze the situation in which the weight for each QI is restricted to a proportional increase above and decrease below the overall prevalence-based weight (this analysis facilitates comparisons with results from the DEA model), using the following proportions: ±0.25, ±0.50, ±0.75, and ±0.90. When showing facility-specific results, we focus on the ±0.75 scenario, which we consider to be a moderate level of weight adjustment. In addition, we solve the model with no constraints. The solution to the unconstrained model is not practical, since all weight is placed on the one QI on which the facility performs the best. However, it does show the largest possible impact from allowing increased levels of weight adjustment.

2.2.3 Second benefit-of-the-doubt approach: DEA

In DEA, organizational units want to maximize efficiency, defined as the ratio of outputs to inputs. When applied to the problem of calculating a composite measure, the outputs are the QIs and the input for each facility is a "dummy variable" set equal to 1. To convert the problem of minimizing a weighted sum of *O/E* ratios (which we used in the simple LP model) to a comparable maximization problem, we maximize the quantity (K − *O/E*), where K is an arbitrary constant. (Note that in the simple LP model, in which the sum of the weights equals 1, maximizing $\Sigma_i w_{if}$ *(K − O_{if}/E_{if}) has the same solution as minimizing $\Sigma_i w_{if}$ (O_{if}/E_{if})). In DEA, the constraint is not on the sum of the weights, but on the composite score, which is constrained to be ≤1. Thus, a facility's composite score represents the proportion of the maximum possible score that the facility has achieved. In the same sense that the optimal weights for a facility in the simple LP model have to be feasible for all other facilities (i.e., they have to sum to 1), the optimal weights for any facility in DEA have to be feasible for all other facilities.

This gives rise to the following decision problem for facility f : find w_{if} to

$$\begin{aligned} &\text{maximize } \Sigma_i w_{if*} \left(\mathrm{K} - O_{if}/E_{if}\right) \\ &\text{subject to } \Sigma_i w_{if*} \left(\mathrm{K} - O_{ij}/E_{ij}\right) \leq 1 \text{ for all } j,\ j = 1, \ldots, 32 \\ &\qquad w_{if} > 0 \text{ for all } i,\ i = 1, \ldots, 5 \end{aligned}$$

Again, we want to ensure that the weights stay within some range of a set of baseline weights. The traditional way to do this in DEA is to constrain the ratio of the weights (Allen et al. 1997). We consider the ratios of the overall prevalence-based weights as baseline. Since in the simple LP model, weights are restricted to the range ± P of overall prevalence-based weights, in the DEA model, the ratio of weights is restricted to the range (1 − P)/(1 + P) to (1 + P)/(1 − P) times the ratio of the overall prevalence-based weights. We also solve the unconstrained model.

To solve the simple LP models efficiently, we use Solver in Excel plus an Excel Add-in, Solver Table (Winston and Albright 2001). To solve the DEA models, we use Excel-based DEA software called NUHOME (which also uses Solver) developed by Crystal Decision Systems, Inc. to enable nursing home managers to benchmark with DEA (Lenard et al. 2004).

3 Results

Table 1 shows the number of residents eligible for each QI, the number who experienced each type of adverse event, the ratio of the number who experienced the adverse event to the number expected to experience the adverse event (the *O/E* ratio), and prevalence-based weights. These data are shown for facilities ranked in the top 6 (high ranked facilities) or bottom 6 (low ranked facilities) by any of the standard approaches or the benefit-of-the-doubt approaches allowing ±0.75 adjustments to overall prevalence-based weights. As shown in Part A of the table, in most facilities, the mortality indicator has the largest number of eligible residents, which ranges from a low of 83 residents to a high of 817 residents (2 facilities not shown in the table also have more than 800 patients). The much lower number of residents eligible for pressure ulcers, functional decline and behavioral decline at some facilities results mainly from lack of a second assessment in many patients with short lengths of stay. Part B shows the number of people experiencing each adverse event. Part C shows the *O/E* ratios. There is wide variation in the *O/E* ratios, though less so for the mortality indicator. Part D shows both facility-specific prevalence-based weights and, in the last row, overall prevalence-based weights. The weight assigned to pressure ulcer development, the lowest prevalence condition, ranges from 0.013 to 0.091; the weight assigned to mortality, the highest prevalence condition, ranges from 0.195 to 0.747. Hence, there is substantial variation in the way in which QIs are weighted when using facility-specific prevalence-based weights. Using overall prevalence-based weights, pressure ulcer development receives relatively little weight (0.044) and mortality a very high weight (0.439).

Table 2 shows composite scores and facility ranks for each of the standard approaches and for the two benefit-of-the-doubt approaches (again with ±0.75 allowable adjustments). Facility-specific prevalence-based weights hurt facilities that do well on quality indicators on which they have a low prevalence (compared to the average prevalence over all facilities). For example, facility 8 does very well on pressure ulcer development, functional decline, and behavioral decline, indicators on which its facility-specific prevalence-based weights are much lower than the overall prevalence-based weights. Equal weighting hurts facilities that do well on high prevalence and poorly on low prevalence indicators (e.g., facility 11, that does particularly well on mortality and poorly on pressure ulcers) and helps facilities that do well on low prevalence indicators (e.g., facility 4). The benefit-of-the-doubt approaches assign more weight to those indicators on which facilities do particular well. This helps facilities that do particularly well on relatively high prevalence indicators (e.g., facility 30 and 32, that do particulary well on mortality, and facility 2 that does particularly well on both mortality and preventable hospitalizations.).

Perhaps most striking from this table is the minimal impact allowing adjustments of ±0.75 has on which facilities are ranked in the top 6. Using DEA, all of the high ranked facilities achieve a composite score of 1, indicating they are potential benchmarks for other facilities. The benefit-of-the-doubt approaches have a much larger effect on which facilities are ranked among the bottom 6. In this case, only 3 of the facilities ranked in the bottom 6 using overall prevalence-based weights are ranked in the bottom 6 using either of the benefit-of-the-doubt approaches. And, the facilities not ranked in the bottom 6 using benefit-of-the-doubt approaches sometimes move up many ranks (e.g., facilities 2, 30 and 32).

Figure 1a shows the effect of the amount of allowable weight adjustment on the average difference in ranks when using overall prevalence-based weights compared to each of the benefit-of-the-doubt weights. Most obvious from this figure is that for a given level of

Table 1 Number of patients eligible for each quality indicator, observed number experiencing each type of adverse event, ratio of observed cases to expected cases, and prevalence-based weights for high and low ranked facilities

Facility	Part A					Part B				
	Number of patients eligible for each quality indicator					Number experiencing each type of adverse event				
	Pressure ulcers	Functional decline	Behavioral decline	Mortality	Preventable hospitalizations	Pressure ulcers	Functional decline	Behavioral decline	Mortality	Preventable hospitalizations
2	92	87	92	177	165	10	30	52	24	11
4	236	240	240	575	545	5	45	22	115	43
6	203	218	142	187	171	3	11	10	28	10
8	35	29	36	278	271	0	3	0	45	22
10	137	141	151	228	213	1	4	5	21	20
11	45	47	49	307	298	4	7	7	81	5
13	244	273	306	644	597	28	44	46	163	66
15	197	208	222	520	505	19	43	46	78	31
16	126	132	160	419	414	8	19	22	112	47
17	159	163	168	376	353	5	18	1	76	28
18	138	133	153	254	236	10	26	40	49	47
19	73	77	85	201	184	3	14	0	40	14
20	203	209	207	229	214	10	64	91	39	11
24	85	93	94	374	360	0	8	7	77	21
28	33	30	37	325	320	0	3	1	103	8
30	85	87	78	83	77	2	34	26	8	9
31	274	248	276	817	810	3	47	37	213	63
32	299	260	305	324	279	7	31	130	42	50

Table 1 continued

Facility	Part C					Part D				
	Ratio of observed cases to expected cases					Facility-specific prevalence-based weights				
	Pressure ulcers	Functional decline	Behavioral decline	Mortality	Preventable hospitalizations	Pressure ulcers	Functional decline	Behavioral decline	Mortality	Preventable hospitalizations
2	2.07	2.26	2.78	0.78	0.69	0.058	0.159	0.223	0.370	0.190
4	0.69	0.96	0.55	1.15	1.22	0.031	0.203	0.174	0.437	0.154
6	0.28	0.37	0.39	0.99	0.44	0.091	0.252	0.220	0.241	0.196
8	0.00	0.54	0.00	0.98	1.47	0.024	0.076	0.069	0.626	0.205
10	0.18	0.15	0.21	0.65	1.35	0.053	0.263	0.227	0.313	0.144
11	1.76	0.79	0.93	0.87	0.23	0.017	0.066	0.056	0.700	0.160
13	1.64	1.17	0.92	1.00	1.26	0.053	0.117	0.157	0.508	0.164
15	1.94	1.17	1.10	0.96	0.79	0.047	0.176	0.200	0.390	0.187
16	1.18	0.84	0.94	0.99	1.51	0.034	0.115	0.119	0.573	0.158
17	0.80	0.55	0.04	1.29	0.96	0.041	0.216	0.159	0.390	0.193
18	1.42	1.08	1.31	1.20	1.59	0.053	0.182	0.232	0.310	0.224
19	0.90	1.10	0.00	0.86	0.92	0.037	0.142	0.133	0.519	0.169
20	1.19	1.64	2.28	1.23	0.58	0.061	0.282	0.289	0.230	0.138
24	0.00	0.57	0.56	0.93	1.04	0.030	0.105	0.094	0.619	0.151
28	0.00	0.60	0.21	1.05	0.37	0.013	0.038	0.036	0.747	0.165
30	0.88	1.71	1.65	0.73	1.21	0.040	0.352	0.279	0.195	0.133
31	0.22	1.29	0.83	1.22	1.23	0.043	0.114	0.139	0.544	0.160
32	0.44	1.10	2.01	0.78	1.48	0.081	0.143	0.329	0.275	0.172
					Overall prevalence-based-weights	0.044	0.172	0.183	0.439	0.162

Table 2 Composite scores and facility ranks for high and low ranked facilities

Facility	Composite score					Facility ranks				
	Overall prevalence-based weights	Facility-specific prevalence-based weights	Equal weights	Simple LP model*	DEA*	Overall prevalence-based weights	Facility-specific prevalence-based weights	Equal weights	Simple LP model*	DEA*
10	0.576	0.495	0.509	0.351	1.000	1	1	3	1	1
6	0.654	0.530	0.495	0.451	1.000	2	2	2	3	1
28	0.662	0.875	0.445	0.412	1.000	3	7	1	2	1
19	0.754	0.790	0.755	0.596	1.000	4	4	8	6	1
8	0.762	0.957	0.599	0.479	1.000	5	15	4	4	1
24	0.779	0.848	0.621	0.636	0.969	6	6	5	7	9
11	0.805	0.782	0.917	0.680	1.000	7	3	16	8	1
17	0.857	0.846	0.728	0.566	0.990	9	5	6	5	8
4	0.997	1.002	0.914	0.868	0.900	21	20	14	22	28
15	1.035	1.037	1.190	0.936	0.939	23	23	27	27	18
16	1.047	1.056	1.091	0.951	0.895	24	24	23	28	29
13	1.086	1.086	1.198	1.001	0.903	26	26	28	29	27
31	1.118	1.134	0.958	1.023	0.881	27	27	18	30	31
30	1.150	1.402	1.236	0.855	0.952	28	30	29	19	14
32	1.158	1.323	1.162	0.858	0.930	29	29	25	20	23
18	1.272	1.303	1.321	1.187	0.837	30	28	30	32	32
20	1.385	1.556	1.383	1.107	0.889	31	32	31	31	30
2	1.443	1.520	1.717	0.923	0.949	32	31	32	25	16

* Results for the benefit-of-the-doubt approaches are for allowable weight adjustments of ±0.75 of overall prevalence-based weights

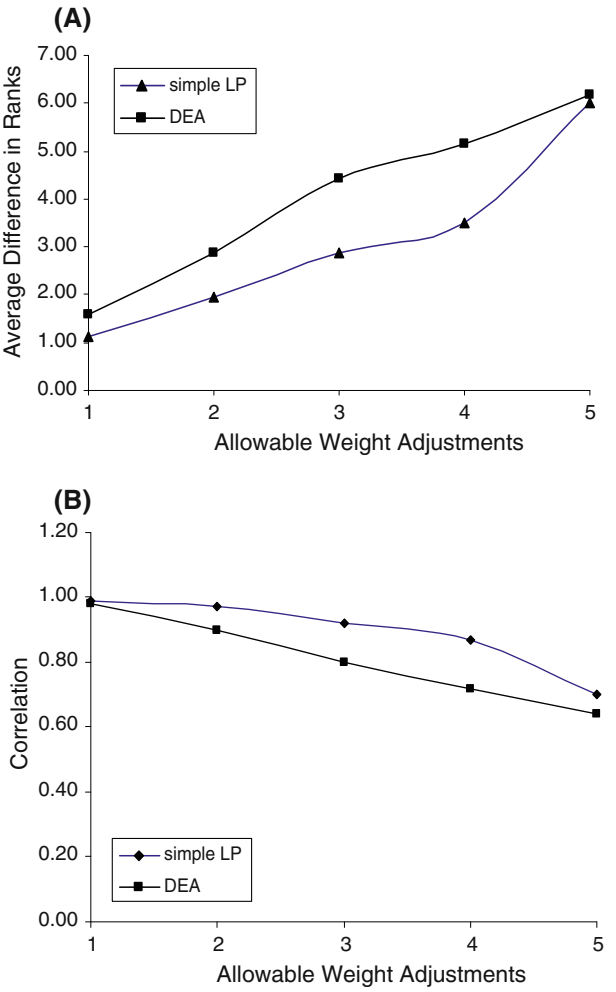


Fig. 1 Comparison of ranks using overall prevalence-based weights to ranks using each of the benefit-of-the-doubt approaches with different amounts of allowable weight adjustments*. Part A: Average difference in ranks. Part B: Correlation. * *Allowable weight adjustments*: 1: $\pm$0.25*overall prevalence-based weights, 2: $\pm$0.50* overall prevalence-based weights, 3: $\pm$0.75* overall prevalence-based weights, 4: $\pm$0.90* overall prevalence-based weights, 5: no constraints

flexibility, DEA has a larger impact on ranks than the simple LP models. For comparison purposes, when using facility-specific prevalence-based weights instead of overall prevalence-based weights, the average change in ranks is 2; when using equal weights instead of overall prevalence-based weights, the average change in ranks is 3. Thus, the simple LP models with allowable adjustments of $\pm$0.75 and the DEA model with a allowable adjustments of $\pm$0.50 would have less of an impact on ranks than switching from overall prevalence-based weights to equal weighting; the simple LP models with allowable adjustments of $\pm$0.50 and the DEA model with allowable adjustments of $\pm$0.25 would have less of an impact on ranks than switching from overall prevalence-based weights to facility-specific prevalence-based weights. At the extreme, when the models are solved without any constraints, the average change in ranks is about 6, which on average would move a facility from one quintile to another.

Figure 1b shows the effect of the amount of allowable weight adjustment on the correlation of ranks when using overall prevalence-based weights compared to each of the

benefit-of-the-doubt weights. Again for comparison purposes, the correlation between ranks using overall prevalence-based weights and facility-specific prevalence-based weights is 0.95; the correlation when using overall prevalence-based weights and equal weights is 0.90. In relation to comparisons with switching between standard approaches, findings are similar to those in Fig. 1. The correlation in ranks when using overall prevalence-based weights compared to equal weighting is no higher than the correlation when using overall prevalence-based weights compared to the simple LP models with allowable adjustments of ± 0.75 or the DEA models with allowable adjustments of ± 0.50. The correlation in ranks when using overall prevalence-based weights compared to facility-specific prevalence-based weights is lower than the correlation when using overall prevalence-based weights compared to the simple LP models with ± 0.50 adjustments or the DEA models with ± 0.25 adjustments.

4 Discussion

There is no gold standard that allows one to say that a composite measure calculated from one set of weights is a better than another. In this situation, the value of any particular approach rests upon its face validity: does it make sense to users? The two most widely used approaches—variants of equal weighting and use of "expert" or "policy maker" judgment—have some appeal. However, it is easy for providers to take "pot shots" at the resulting weights, often expressed by statements like the following: "These weights do not reflect what is important to our patients."

Since benefit-of-the-doubt weights apply to retrospective data, the above statement does reflect a self-serving tendency noted by Gormley and Weimer (1999) for organizations to "argue for favorable weights once they know their own position on the dependent variables underlying the scales." However, to the extent a facility's "position on the dependent variables" reflects its preferences about the relative importance of the dependent variables, there is a legitimate basis for this argument. Benefit-of-the-doubt approaches, by giving facilities the benefit-of-the-doubt about the reasons for differences in relative performance across QIs, recognize this legitimacy and thus should lessen tensions between policy makers and providers. In particular, these approaches should reduce the tendency noted by Gormley and Weimer (1999) for poorly ranked facilities to "blame the messenger." If this does happen, the messenger has a reasonable response: "Look, within the specified levels of allowable weight adjustments, we let you weight those QIs on which you performed well higher and those on which you did not perform well lower. Even after letting you do this, you still did not perform well compared to other facilities." Facilities might argue that the level of weight adjustments is not sufficient, but this has the potential to be a more productive conversation than one criticizing the fairness of the baseline weights.

There are a variety of legitimate reasons specific to providers' local context that might partially explain differences in relative performance on the QIs. For example, risk adjustment is not perfect. Poor performance on the mortality QI may reflect a particularly severe resident case mix that is not adequately captured by variables in the risk adjustment model. High rates of preventable hospitalizations (which can be reduced if there is adequate physician availability at a nursing home) may reflect an isolated rural facility with little physician availability in the community. For many QIs, the link between processes used in resident care and resident outcomes is weak. For example, for an indicator like functional decline, a particular facility may not believe that there is sufficient evidence that process improvements can lead to better outcomes among its particularly frail patient

population, and hence may not focus on this indicator even though its performance is poor. It is these types of factors that are taken into account by the allowable weights adjustments with benefit-of-the-doubt approaches. To the extent policy makers feel that factors like these exist and are important, more flexibility can be allowed in the weight adjustments; to the extent they do not, less flexibility can be allowed. And, to the extent they feel "legitimate" factors exist for some indicators but less so for others, different levels of allowable weight adjustments could be specified for different indicators.

It is interesting in our example that benefit-of-the-doubt approaches with moderate levels of allowable weight adjustment (e.g., ± 0.75) had little impact on which facilities were identified as top performers, but a larger impact on which providers were identified as poor performers. There is no particular reason to think this result generalizes. The more important lesson from our example is the benefit of examining the relationship between amount of allowable weight adjustment and impact on rankings. Understanding this relationship would be an important factor in deciding upon the amount of flexibility to allow when implementing a benefit-of-the-doubt approach.

We considered two benefit-of-the-doubt approaches. To the best of our knowledge, DEA is the only one of these approaches that has been used by others. As noted, one advantage of DEA is that quality indicators measured using different metrics can be combined without first normalizing the individual scores. In addition, DEA identifies for each facility performing below its potential, benchmark facilities (or a benchmark facility). The composite score has a clear interpretation: it is the percentage of a facility's potential that is being realized. Finally, benefit-of-the-doubt weights obtained through DEA have nice theoretical properties: as noted, they can be conceived of as Nash equilibria in a game theoretic formulation. However, there are some disadvantages as well. When the number of facilities is relatively small (as it is in our case), it can be difficult to reliably identify appropriate benchmark facilities. Also, if facility-specific prevalence-based weights were used as the baseline weights, the constraints on the ratio of weights would be different for each facility. This type of situation has not been considered in the DEA literature to the best of our knowledge and the implications for the efficiency frontier are not clear. In addition, and most important, it is difficult to determine exactly what it is about a facility's performance that results in the specific weights identified and the resulting composite score. The reason is that a facility's weights are determined not only by its own performance but by the performance of all other facilities (since the weights selected for a particular facility have to be feasible for all other facilities). The simple linear programming model approach can not handle indicators measured on different scale without first normalizing the scores. However, it has one big advantage: its transparency. It is easy to justify the resulting weights to policy makers and facility managers.

An important question is "when do facilities benefit from benefit-of-the doubt approaches?" The grading example discussed in the Introduction provides some insight. The student who has relatively similar scores on all 5 assignments will not receive value from being able to drop the lowest assignment and, unless all 5 scores are near the top, is likely to be negatively impacted by a policy that allows students to drop their lowest assignment. In the same sense, a facility that performs relatively similarly across indicators will not receive advantage from benefit-of-the-doubt weights and, in terms of rank relative to other facilities, is likely to suffer. Thus, benefit-of-the doubt approaches encourage managers to proactively consider which indicators are most important for their patients and which they believe they can most favorably impact, and to then focus on these QIs; or, alternatively, to identify those indicators that are less important for their patients, or particularly difficult or costly to change, and to de-emphasize these indicators.

Many pay-for-performance programs reward facilities both on the level of quality indicators at the end of the reporting period and on the amount of improvement in the indicators over the reporting period. Benefit-of-the-doubt approaches could be used to calculate a composite measure separately for absolute levels of the indicators and for changes in the indicators. These two composites could then be weighted using whatever weights policy makers decide upon. Or, alternatively, both absolute and change measures could be simultaneously weighted using benefit-of-the-doubt approaches.

We have illustrated our approach by making adjustment to baseline weights. However, there are a variety of alternative ways in which limitations might be placed on the weights. For example, policy makers might specify a rank ordering of the weights but then allow any specific weights consistent with this rank ordering; or, they might specify that the weight assigned to a particular indicator has to be 25% greater than the weight assigned to any other indicator. A more formal consensus process that uses modified Delphi or other similar methods to achieve consensus among stakeholders might be used to determine constraints on the weights. This is an example of a hybrid approach that combines expert judgment with benefit-of-the-doubt approaches.

Other forms of the outcome measures can be used. One disadvantage of *O/E* ratios is that they do not take account of differences in reliability of estimates from facilities of different sizes. As an alternative, one might consider z scores, e.g., $(O - E)$/variance (O). An advantage of converting a measure to a z score is that measures with very different scales can be easily combined. Or, instead of the observed performance on each indicator, one might use shrunken estimates of performance from hierarchical models (Arling et al. 2007).

We have illustrated benefit-of-the-doubt approaches for calculating a composite measure from individual indicators. However, the principle underlying the approach—that relative performance on a set of indicators reflects revealed preferences about the relative importance of the indicators—may have some value in the design phase of composite measures. CMS recently starting to post on its website results from the *Nursing Home Compare* Five-Star Quality Rating System. There is a fairly complex algorithm that leads to the final assignment of one to five stars to each nursing home. The people who contributed to the creation of the algorithm no doubt thought carefully about the relative importance of various components of the composite and developed an approach for arriving at a composite that reflected these judgments. In its current form, the algorithm does not lend itself to weight adjustments. As an additional consideration in the design phase, designers could have considered the question "Where might reasonable tradeoffs be made by nursing homes concerning elements of the composite?" This might have led to a different approach to calculating the composite, one that allowed some adjustments in weightings for those components of the composite about which facilities might reasonably make tradeoffs.

We studied only 32 nursing homes in a particular environment, the Veterans Health Administration. However, our purpose in this paper was not to undertake an actual evaluation. Rather, it was to use these data as an example to illustrate a potentially useful modification of existing ways that have been proposed for weighting individual QIs when calculating a composite score. In an environment like the Veterans Health Administration, nursing home chains, or perhaps in states, where policy makers desire to work collaboratively with providers, benefit-of-the-doubt approaches seem an option worth considering. These approaches should be attractive to facilities and may help reduce tensions between policy makers and facilities, which each have to live with the often implicit priorities imbedded in decisions about weights. Under benefit-of-the-doubt approaches, providers

who invest significant resources in achieving high performance on a particular set of QIs are assured baseline weights will be adjusted to reflect this success, as are other providers who focus on a different set of QIs. Further, as providers and policy makers better understand the implications of different amounts and different types of flexibility in choosing weights, an opportunity is created for healthy dialogue concerning future adaptations of the "rules of the game." Such dialogue can only improve the value of composite measures and perhaps increase trust and good will that may translate into better societal outcomes.

Acknowledgments This study was supported by VA Health Services Research and Development Service grant IIR 06-260. Findings and conclusions do not necessarily reflect the opinions or policies of the Department of Veterans Affairs.

References

Agency for Healthcare Research and Quality: Prevention quality indicators overview. http://www.qualityindicators.ahrq.gov/pqi_overview.htm (2004). Accessed 4 Nov 2008

Agency for Healthcare Research and Quality (AHRQ) Quality Indicators: Patient Safety Indicator Composite Measure Final Technical Report (October 2006)

Allen, R., Athanassopoulos, A., Dyson, R.G., Thanassoulis, E.: Weights restrictions and value judgements in data envelopment analysis. Ann. Oper. Res. **73**, 13–34 (1997)

Arling, G., Karon, S.L., Sainfort, F., Zimmerman, D.R., Ross, R.: Risk adjustment of nursing home quality indicators. Gerontologist **37**, 757–766 (1997)

Arling, G., Lewis, T., Kane, R.L., Mueller, C., Flood, S.: Improving quality assessment through multilevel modeling: the case of nursing home compare. Health Serv. Res. **42**, 1177–1199 (2007)

Ash, A.S., Shwartz, M., Peköz, E.: Comparing outcomes across providers. In: Iezzoni, L.I. (ed.) Risk Adjustment for Measuring Health Care Outcomes, 3rd edn, pp. 297–333. Health Administration Press, Chicago (2003)

Berlowitz, D.B., Ash, A.S., Brandeis, G.H., Brand, H.K., Halpern, J.L., Moskowitz, M.A.: Rating long-term care facilities on pressure ulcer development: importance of case-mix adjustment. Ann. Intern. Med. **124**, 557–563 (1996)

Berlowitz, D.R., Young, G.J., Hickey, E.C., Saliba, D., Mittman, B.S., Czarnowski, E., et al.: Quality improvement implementation in the nursing home. Health Serv. Res. **38**, 65–83 (2003)

Berlowitz, D.B., Rosen, A.K., Wang, F., Tsilimingras, D., Tariot, P.N., Engelhardt, B., et al.: Purchasing or providing nursing home care: can quality of care data provide guidance. J. Am. Geriatr. Soc. **53**, 603–608 (2005)

Braun, B.I.: The effect of nursing home quality on patient outcome. J. Am. Geriatr. Soc. **39**, 329–338 (1991)

Caldis, T.: Composite health plan quality scales. Health Care Financ. Rev. **28**, 95–107 (2007)

Carter, M.W.: Factors associated with ambulatory care-sensitive hospitalizations among nursing home residents. J Aging Health **15**, 295–330 (2003)

Charnes, A., Cooper, W.W., Rhodes, E.: Measuring the efficiency of decision making units. Eur. J. Oper. Res. **2**, 429–444 (1978)

Cherchye, L., Kuosmanen, T.: Benchmarking sustainable development: a synthetic meta-index approach. In: McGillivray, M., Clarke, M. (eds.) Perspectives on human development, chapter 7. United Nations University Press, Tokyo (2004)

Cherchye, L., Moesen, W., Rogge, N., Van Puyenbroeck, T.: An introduction to 'benefit of the doubt' composite indicators. Soc. Indic. Res. **82**, 111–145 (2007)

Cherchye, L., Moesen, W., Rogge, N., Van Puyenbroeck, T., Saisana, M., Saltelli, A., Liska, R., Tarantola, S.: Creating composite indicators with DEA and robustness analysis: the case of the Technology Achievement Index. J. Oper. Res. Soc. **59**, 239–251 (2008)

Cooper, W.W., Seiford, L.M., Tone, K.: Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software. Springer Science and Business Media, LLC, New York (2007)

Despotis, D.K.: A reassessment of the Human Development Index via data envelopment analysis. J. Oper. Res. Soc. **56**, 969–980 (2005)

Edwards, J.R., Bagozzi, R.P.: On the nature and direction of relationships between constructs and measures. Psychol. Methods **5**, 155–174 (2000)
Feinstein, A.R.: Multi-item "instruments" vs Virginia Apgar's principles of clinimetrics. Arch. Intern. Med. **159**, 125–128 (1999)
Gandhi, T.K., Cook, E.F., Puopolo, A.L., Burstin, H.R., Haas, J.S., Brennan, T.A.: Inconsistent report cards: assessing the comparability of various measures of the quality of ambulatory care. Med. Care **40**, 155–165 (2002)
Gormley, W.T., Weimer, D.L.: Organizational Report Cards. Harvard University Press, Cambridge MA (1999)
Institute of Medicine: Performance Measurement: Accelerating Improvement. National Academy Press, Washington, DC (2006)
Jacobs, R., Goddard, M., Smith, P.C.: How robust are hospital ranks based on composite performance measures. Med. Care **43**, 1177–1184 (2005)
Jarvis, C.B., MacKenzie, S.B., Podsakoff, P.M.: A critical review of construct indicators and measurement model misspecification in marketing and consumer research. J. Consum. Res. **30**, 199–218 (2003)
Jha, A.K., Zhonghe, L., Orav, E.J., Epstein, A.M.: Care in U.S. hospitals—The Hospital Quality Alliance Program. N. Engl. J. Med. **353**, 265–274 (2005)
Landrum, M.B., Bronskill, S.E., Normand, S.-L.: Analytic methods for constructing cross-sectional profiles of health care providers. Health Serv. Outcomes Res. Methodol. **1**, 23–47 (2000)
Lenard, M.L., Wagner, J.M., Shimshak, D.G., Porell, F.W., Klimberg, R.K.: Evaluating the performance of nursing homes using data envelopment analysis. In: Lawrence, K.D. (ed.) Mathematical Programming, Applications of Management Science, vol. 11, pp. 89–105. Elsevier Ltd, Amsterdam (2004)
Lied, T.R., Malsbary, R., Eisenberg, C., Ranck, J.: Combining HEDIS indicators: A new approach to measuring plan performance. Health Care Financ. Rev. **23**, 117–129 (2002)
Lindenauer, P.K., Remus, D., Roman, S., Rothberg, M.B., Benjamin, E.M., Ma, A., et al.: Public reporting and pay for performance in hospital quality improvement. N. Engl. J. Med. **356**, 486–496 (2007)
Lovell, C.A.K.: Measuring the macroeconomic performance of the Taiwanese economy. Int. J. Prod. Econ. **39**, 165–178 (1995)
Mahlberg, B., Obersteiner, M.: Remeasuring the HDI by data envelopment analysis. Interim Report IR-01-069. International Institute for Applied Systems Analysis, Laxenburg, Austria (2001)
Mukamel, D.B.: Risk-adjusted outcome measures and quality of care in nursing homes. Med. Care **35**, 367–385 (1997)
Nardo, M., Saisana, M., Saltelli, A., Tarantola S. Handbook on constructing composite indicators: methodology and user guide. Organization for Economic Co-operation and Development (OECD) Statistics Working Paper, 2005. http://www.olis.oecd.org/olis/2005doc.nsf/LinkTo/std-doc(2005)3. Accessed 9 Sept 2009
O'Brien, S.M., Shahian, D.M., DeLong, E.R., Normand, S.-L.T., Edwards, F.H., Ferraris, V.A., et al.: Quality measurement in adult cardiac surgery: part 2—statistical considerations in composite measure scoring and provider rating. Ann. Thorac. Surg. **83**, S13–S26 (2007)
Pizer, S.D., Wang, M., Comstock, C.: Preventable hospitalization as a measure of quality of care in nursing homes. Working Paper #2003-01. Health Care Finance and Economics, VA Bedford, MA (2003)
Porell, F., Caro, F.G., Silva, A., Monane, M.: A longitudinal analysis of nursing home outcomes. Health Serv. Res. **33**, 835–865 (1998)
Premier: Hospital Quality Incentive Demonstration Project. Summary of Composite Quality Scoring Methodology. http://www.premierinc.com/quality-safety/tools-services/p4p/hqi/resources/top-performer-summary.pdf. (2003). Accessed 4 Nov 2008
Reeves, D., Campbell, S.M., Adams, J., Shekelle, P.G., Kontopantelis, E.: Combining multiple indicators of clinical quality: an evaluation of different analytic approaches. Med. Care **45**, 489–496 (2007)
Rosen, A., Wu, J., Chang, B., Berlowitz, D.R., Rakovski, C., Ash, A.S., et al.: Risk adjustment for measuring health outcomes: an application in VA long-term care. Am. J. Med. Qual. **16**, 118–127 (2001)
Ryan, A.M., Burgess, J.F., Tompkins, C.P., Wallack, S.S.: The relationship between Medicare's process of care quality measures and mortality. Inquiry **46**, 274–291 (2009)
Semple, J.: Constrained games for evaluating organizational performance. Eur. J. Oper. Res. **96**, 103–112 (1996)
Shwartz, M., Ren, J., Peköz, E.A., Wang, X., Cohen, A.B., Restuccia, J.D.: Estimating a composite measure of hospital quality from the Hospital Compare database: differences when using a Bayesian hierarchical latent variable model versus denominator-based weights. Med. Care **46**, 778–785 (2008)
Werner, R.M., Bradlow, E.T.: Relationship between Medicare's hospital compare performance measures and mortality rates. JAMA **296**, 2694–2702 (2006)
Winston, W., Albright, S.C.: Practical Management Science, 2nd edn. Duxbury, Pacific Grove, CA (2001)

Zaslavsky, A.M., Shaul, J.A., Zaborski, L.B., Cioffi, M.J., Cleary, P.D.: Combining health plan performance indicators into simpler composite measures. Health Care Financ. Rev. **23**, 101–115 (2002)
Zimmerman, D.R., Karon, S.L., Arling, G., Clark, B.R., Collins, T., Ross, R., et al.: Development and testing of nursing home quality indicators. Health Care Financ. Rev. **16**, 107–127 (1995)

A Single-Question Screening Test for Drug Use in Primary Care

Peter C. Smith, MD, MSc; Susan M. Schmidt, BA; Donald Allensworth-Davies, MSc; Richard Saitz, MD, MPH

Background: Drug use (illicit drug use and nonmedical use of prescription drugs) is common but underrecognized in primary care settings. We validated a single-question screening test for drug use and drug use disorders in primary care.

Methods: Adult patients recruited from primary care waiting rooms were asked the single screening question, "How many times in the past year have you used an illegal drug or used a prescription medication for nonmedical reasons?" A response of at least 1 time was considered positive for drug use. They were also asked the 10-item Drug Abuse Screening Test (DAST-10). The reference standard was the presence or absence of current (past year) drug use or a drug use disorder (abuse or dependence) as determined by a standardized diagnostic interview. Drug use was also determined by oral fluid testing for common drugs of abuse.

Results: Of 394 eligible primary care patients, 286 (73%) completed the interview. The single screening question was 100% sensitive (95% confidence interval [CI], 90.6%-100%) and 73.5% specific (95% CI, 67.7%-78.6%) for the detection of a drug use disorder. It was less sensitive for the detection of self-reported current drug use (92.9%; 95% CI, 86.1%-96.5%) and drug use detected by oral fluid testing or self-report (81.8%; 95% CI, 72.5%-88.5%). Test characteristics were similar to those of the DAST-10 and were affected very little by participant demographic characteristics.

Conclusion: The single screening question accurately identified drug use in this sample of primary care patients, supporting the usefulness of this brief screen in primary care.

Arch Intern Med. 2010;170(13):1155-1160

Author Affiliations: Section of General Internal Medicine, Department of Medicine (Drs Smith and Saitz and Ms Schmidt), and Clinical Addiction Research and Education (CARE) Unit (Dr Saitz), Boston Medical Center and Boston University School of Medicine, and Data Coordinating Center (Mr Allensworth-Davies) and Youth Alcohol Prevention Center and Department of Epidemiology (Dr Saitz), Boston University School of Public Health, Boston, Massachusetts.

ILLICIT DRUG USE AND NONMEDICAL use of prescription drugs are common in the primary care setting and are underrecognized.[1,2] Screening for drug use allows clinicians to counsel patients and, when indicated, refer them to treatment. Because of this, the Substance Abuse and Mental Health Services Administration has promoted the integration of screening and brief intervention for substance use disorders into the primary care setting.[3] Screening for drug use is also useful as part of routine clinical care, for instance to aid in diagnosis and to avoid medication interactions. Few screening instruments for drug use or drug disorders have been validated, however, for use in primary care settings. Time is also limited during the primary care office visit, and commonly recommended drug screening instruments are composed of multiple questions, can be time consuming to administer, and may require scoring.[4,5] Practice guidelines currently recommend the use of a single screening question for the detection of unhealthy alcohol use in primary care settings.[6] Analogous single screening questions may also improve screening for drug use. We therefore set out to validate such a screening question in a sample of primary care patients.

METHODS

PARTICIPANTS

The study was conducted between October 2006 and June 2007 at an urban safety-net hospital-based primary care clinic at an academic medical center. The participant selection and data collection methods have been described previously.[7] Briefly, a sample of patients in the waiting room was selected by a research associate who systematically approached those waiting to be seen according to a predetermined pattern based on waiting room seating, which was varied daily. This was done to minimize biased selection of participants, because, owing to the large number of patients attending the clinic, all patients could not be approached. Prior to being approached for eligibility screening, patients saw no advertisement or indication by the research associate as to what the study was about. Patients who were younger than 18 years were excluded, as were those who, in the judgment of the research associate, would be unable to complete the questionnaire because of limited English, cognitive impairment, or acute illness. People in the waiting room accompanying patients who reported that they themselves were not patients of the clinic were also excluded. The institutional review board of Boston University Medical Center, Boston, Massachusetts, reviewed and approved all study procedures.

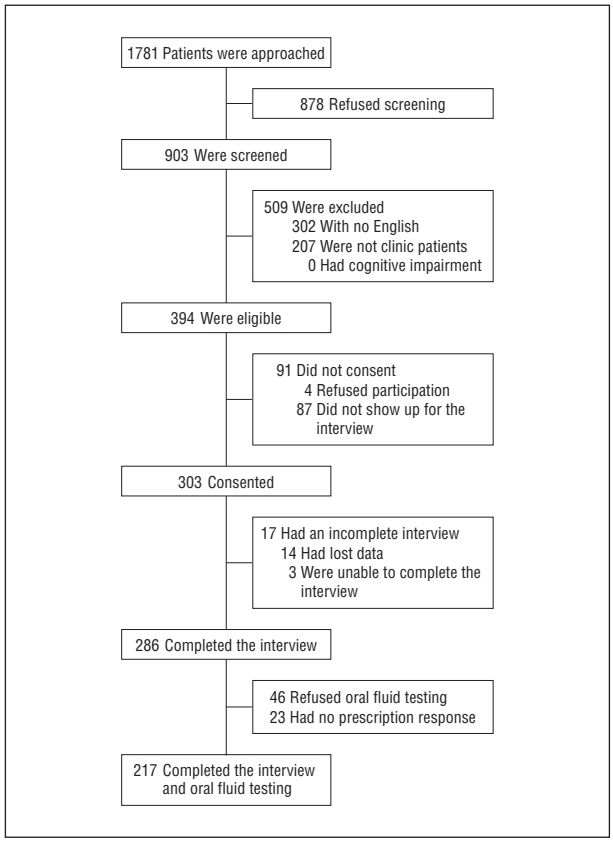


Figure. Flowchart of participant recruitment.

DATA COLLECTION

Interviews were conducted by trained research staff in a private setting, and data were recorded anonymously, unaccompanied by any unique identifiers. Participants were first asked the single screening question, "How many times in the past year have you used an illegal drug or used a prescription medication for nonmedical reasons?" (where a response of ≥1 time was considered positive for drug use). If asked to clarify the meaning of "nonmedical reasons," the research associate added "for instance because of the experience or feeling it caused." After participants responded to the single screening question, they were asked if they had ever experienced any of a list of problems related to drug use. For this we modified the previously described Short Inventory of Problems–Alcohol and Drug (SIP-AD) questionnaire, which asks about problems ever experienced in the participant's lifetime related to alcohol or drug use.[8] We modified this by eliminating the word alcohol from the questions, a modification we hereafter refer to as the Short Inventory of Problems–Drug Use (SIP-DU). In a separate analysis (but in these participants) we determined the reliability and validity of the SIP-DU as a measure of drug use consequences.[9] The computerized version of the Composite International Diagnostic Interview (CIDI) Substance Abuse Module was used for the assessment of current (12-month) drug use disorders.[10] This structured interview yields a *Diagnostic and Statistical Manual of Mental Disorders (Fourth Edition)* diagnosis of drug abuse or dependence. In addition, as part of the CIDI, individuals were asked detailed questions about current (past year) use of illicit drugs (marijuana, cocaine, heroin, stimulants, or hallucinogens) and nonmedical use of prescription drugs. Following the interview, participants were asked to undergo oral fluid testing for the presence of common drugs of abuse (opiates, benzodiazepines, cocaine, methamphetamines, or tetrahydrocannabinol). Once collected, oral fluid was sent to an outside laboratory for analysis using methods that yielded results comparable to urine drug screening (Intercept immunoassay; OraSure Technologies, Bethlehem, Pennsylvania).[11-14] To aid in the interpretation of drug test results, individuals had been asked, as part of the interview, if they had recently been prescribed any drugs from a list of opiates or benzodiazepines. Because this question was added to the questionnaire during the study, responses were missing from 23 patients who underwent oral fluid testing. Participants were not told that they would be asked to undergo drug testing until the interview was complete. After completing the interview, they were compensated and thanked for their participation. They were then asked to undergo oral fluid testing, and a second informed consent process was completed. Following the single drug screening question, but before the other assessments, the 10-item Drug Abuse Screening Test (DAST-10) was administered for comparison.[4] As part of a parallel study on screening for unhealthy alcohol use, participants were also asked a single alcohol screening question (preceding the drug screening question), 2 other brief alcohol screening questionnaires, and a calendar-based assessment of past-month alcohol consumption (all after the drug screen and prior to the CIDI).[7]

REFERENCE STANDARD

Participants were considered to have current drug use if, during the CIDI, they reported the use of an illicit drug (marijuana, cocaine, heroin, stimulants, or hallucinogens), or the use of a prescription drug for nonmedical reasons, during the past 12 months. A second analysis included only individuals who consented to oral fluid testing. Participants in this analysis were considered to have current drug use if they met these criteria; if oral fluid testing was positive for cocaine, tetrahydrocannabinol, or methamphetamines; or if it was positive for opiates or benzodiazepines and they had not reported receiving a recent prescription for one of these medications. Participants were considered to have drug-related problems if they had current drug use and responded positively to any of the 15 SIP-DU questions. Those with drug abuse or dependence as determined by the CIDI and who reported experiencing symptoms within the past 12 months were considered to have a current drug use disorder.

STATISTICAL ANALYSIS

We calculated the sensitivity, specificity, likelihood ratios, and area under the receiver operating curve (AUC) of the single-question screen for the detection of drug use, drug use associated with problems, and a current drug use disorder as defined in the previous subsection. The AUC, a measure of a test's discriminatory power, can be interpreted as the probability, given 1 participant without drug use and 1 with drug use drawn at random from the population, that the person with drug use will score higher on the test. An AUC of 1.0 indicates perfect discrimination, an AUC higher than 0.8 indicates good discrimination, and an AUC of less than 0.7 indicates poor discrimination.[15] For comparison with the single-question screen, we calculated the sensitivity, specificity, likelihood ratios, and AUC of another longer screening test, the DAST-10, for the detection of the same conditions. The DAST-10 yields a score of 0 to 10. A total of more than 2 points is considered a positive screening test result.[4] We calculated 95% confidence intervals (CIs) using published formulas.[16] Statistical analyses were performed using SAS software (version 9.1; SAS Institute Inc, Cary North Carolina).

RESULTS

PARTICIPANT RECRUITMENT

Of the 1781 people approached, 903 (51%) agreed to be screened for study eligibility (**Figure**). Of these, 509 (56%)

Table 1. Participant Characteristics

Characteristic	No. (%)	
	Participants Consenting to Oral Fluid Testing (n=240)	Total (n=286)
Sex		
Female	135 (56.2)	155 (54.2)
Age, mean (SD), y	49.3 (12.8)	49.0 (12.3)
Median (range), y	49.0 (21.0-86.0)	49.0 (21.0-86.0)
Education		
Some high school	68 (28.4)	81 (28.3)
High school graduate	86 (35.8)	107 (37.4)
Some college	50 (20.8)	59 (20.6)
College graduate	26 (10.8)	28 (9.8)
Postgraduate education	10 (4.2)	11 (3.9)
Race/ethnicity		
American Indian/Alaskan native	5 (2.1)	8 (2.8)
Asian	5 (2.1)	7 (2.4)
Black or African American	153 (63.8)	179 (62.6)
Native Hawaiian/Pacific Islander	2 (0.8)	3 (1.1)
White	42 (17.4)	49 (17.1)
Unknown	33 (13.8)	40 (14.0)
Hispanic or Latino ethnicity	38 (15.8)	46 (16.1)
English is first language	185 (77.1)	223 (78.0)
Alcohol use		
Hazardous consumption amounts[a]	71 (29.6)	88 (30.8)
Any lifetime alcohol use disorder (abuse or dependence)[b]	106 (44.2)	126 (44.1)
Drug use[c]		
Current use (self-reported)[d]	86 (35.8)	99 (34.6)
Current use without drug-related problems[e]	6 (2.5)	7 (2.4)
Problem use (current use and drug problem or drug use disorder)	80 (33.3)	92 (32.2)
Current use (either self-reported or a positive oral fluid test)[d]	97 (40.4)	110 (38.5)
Current use without drug-related problems[e]	6 (2.5)	7 (2.4)
Problem use (current use and drug problem or drug use disorder)	91 (37.9)	103 (36.1)
Did not self-report current use	11 (4.6)	11 (3.8)
Current (12 mo) drug abuse[b]	3 (1.2)	3 (1.0)
Current drug dependence[b]	27 (11.2)	34 (11.9)
Any lifetime drug-related problem[f]	118 (49.2)	137 (47.9)
Any lifetime drug use disorder (either abuse or dependence)[b]	116 (48.3)	133 (46.5)
Oral fluid testing		
Any positive test result	44 (18.3)	44 (15.4)
Cocaine	25 (10.4)	25 (8.7)
Methamphetamine	0	0
Tetrahydrocannabinol	8 (3.3)	8 (2.8)
Illicit drug (cocaine, methamphetamine, or tetrahydrocannabinol)	33 (13.8)	33 (11.5)
Opiates		
Reported no prescription	5 (2.1)	5 (1.7)
Reported prescription	5 (2.1)	5 (1.7)
Missing prescription response	2 (0.8)	2 (0.7)
Benzodiazepines		
Reported no prescription	4 (1.7)	4 (1.4)
Reported prescription	4 (1.7)	4 (1.4)
Missing prescription response	1 (0.4)	1 (0.3)

[a] For men, an average of more than 14 drinks per week over the past 30 days, or more than 4 drinks on any 1 day during the past 30 days (for women, >7 drinks per week, or >3 drinks per occasion), determined using a calendar-based reporting method.
[b] Lifetime and current alcohol and drug use disorders as determined by responses to the Composite International Diagnostic Interview (CIDI).
[c] "Current" refers to within the past year (12 months).
[d] As part of the CIDI interview participants are asked about their current use of illicit drugs or of prescription drugs for nonmedical reasons.
[e] Participants were considered to have drug-related problems if they were past-year drug users and responded positively to any of the 15 Short Inventory of Problems-Drug Use (SIP-DU) questions.
[f] A positive response to any of the questions from the SIP-DU questionnaire.

were ineligible for the study: 302 (33%) did not speak English and 207 (23%) were not clinic patients. Of the 394 patients who were eligible, 303 (76%) participated: 4 (1%) refused to participate, 87 (22%) did not show up for the planned interview after the visit with their physician, and of the 303 individuals who arrived and gave consent to participate, 3 (1%) were unable to complete the interview. The data of 14 participants (5%) were lost owing to an electronic error, leaving 286 whose data were analyzed (73% of those eligible).

After completion of the interview, patients were asked to undergo oral fluid testing for common drugs of abuse, to which 240 (84%) consented. Of these, 217 were asked about a recent prescription for opiates or benzodiazepines.

PARTICIPANT CHARACTERISTICS

Of the 286 participants, 54% were women, and the median age was 49 years (range, 21-86 years) (**Table 1**). Most

Table 2. Sensitivity, Specificity, and Likelihood Ratios for the Detection of Drug Use: Single Screening Question

Detection	Sensitivity, % (95% CI)	Specificity, % (95% CI)	Positive LR,[a] (95% CI)	Negative LR,[b] (95% CI)	AUC
Current use, self-reported (n=286)[c]	92.9 (86.1-96.5)	94.1 (89.8-96.7)	15.8 (8.9-28.1)	0.08 (0.04-0.2)	0.93
With drug problem or drug use disorder[d]	93.5 (86.5-97.0)	91.2 (86.4-94.5)	10.7 (6.8-16.8)	0.07 (0.03-0.2)	0.90
Current use, either self-reported or a positive oral fluid test result (n=217)[c]	84.7 (75.6-90.8)	96.2 (91.4-98.4)	22.4 (9.4-53.1)	0.2 (0.1-0.3)	0.92
With drug problem or drug use disorder[e]	84.8 (75.3-91.1)	92.8 (87.2-96.0)	11.7 (6.4-21.4)	0.2 (0.1-0.3)	0.89
Current drug use disorder (n=286)[c]	100 (90.6-100)	73.5 (67.7-78.6)	3.8 (3.1-4.6)	NC	NC

Abbreviations: AUC, area under the receiver operating curve; CI, confidence interval; LR, likelihood ratio; NC, number could not be calculated (ie, the formula results in division by 0).
[a]Calculated as the probability of an individual with the condition having a positive test divided by the probability of an individual without the condition having a positive test.
[b]Calculated as the probability of an individual with the condition having a negative test divided by the probability of an individual without the condition having a negative test.
[c]"Current" refers to within the past year (12 months).
[d]Self-reported past-year drug use and either a positive response to one of the Short Inventory of Problems-Drug Use (SIP-DU) questions or a current drug use disorder.
[e]Self-reported past-year drug use or a positive oral fluid test and either a positive response to one of the SIP-DU questions or a current drug use disorder (excludes participants without an oral fluid test).

Table 3. Sensitivity, Specificity, and Likelihood Ratios for the Detection of Drug Use: DAST-10 Findings in 286 Participants

Detection	Sensitivity, % (95% CI)	Specificity, % (95% CI)	Positive LR,[a] (95% CI)	Negative LR,[b] (95% CI)	AUC
Current use, self-reported (n=286)[c]	82.8 (74.2-89.0)	93.6 (89.1-96.3)	12.9 (7.4-22.5)	0.2 (0.1-0.3)	0.89
With drug problem or drug use disorder[d]	87.0 (78.6-92.4)	92.8 (88.2-95.6)	12.0 (7.2-20.1)	0.1 (0.08-0.2)	0.88
Current use, either self-reported or a positive oral fluid test result (n=217)[c]	80.0 (70.3-87.1)	93.9 (88.5-96.9)	13.2 (6.7-26.0)	0.2 (0.1-0.3)	0.89
With drug problem or drug use disorder[e]	83.5 (73.8-90.1)	92.8 (87.2-96.0)	11.5 (6.3-21.1)	0.2 (0.1-0.3)	0.89
Current drug use disorder (n=286)[c]	100 (90.6-100)	77.1 (71.5-81.9)	4.4 (3.5-5.5)	NC	NC

Abbreviations: AUC, area under the receiver operating curve; CI, confidence interval; DAST-10, 10-item Drug Abuse Screening Test; LR, likelihood ratio; NC, number could not be calculated (ie, the formula results in division by 0).
[a]Calculated as the probability of an individual with the condition having a positive test divided by the probability of an individual without the condition having a positive test.
[b]Calculated as the probability of an individual with the condition having a negative test divided by the probability of an individual without the condition having a negative test.
[c]"Current" refers to within the past year (12 months).
[d]Self-reported past-year drug use and either a positive response to one of the SIP-DU questions or a current drug use disorder.
[e]Self-reported past-year drug use or a positive oral fluid test and either a positive response to one of the SIP-DU questions or a current drug use disorder (excludes participants without an oral fluid test).

participants (63%) identified themselves as black or African American, with whites (17%) and Hispanics (16%) comprising most of the remainder. Most (78%) had completed high school, but only 14% had completed college. The prevalence of self-reported current (past-year) drug use was 35% (with 32% reporting at least 1 problem relating to use), and among those who consented to oral fluid testing, 40% either self-reported drug use or had a positive test result (38% with problem use). The prevalence of current drug abuse or dependence was 13%. The lifetime prevalence of alcohol use disorders (44%) and drug use disorders (47%) was high.

TEST CHARACTERISTICS

The single-question screen was 100% sensitive (95% CI, 90.6%-100%) and 73.5% specific (95% CI, 67.7%-78.6%) for the detection of a current drug use disorder (**Table 2**). It was slightly less sensitive (92.9%; 95% CI, 86.1%-96.5%) and was more specific (94.1%; 95% CI, 89.8%-96.7%) for the detection of current drug use (although CIs overlapped). If oral fluid test results were taken into account, the sensitivity for detecting current drug use was lower (84.7%; 95% CI, 75.6%-90.8%). The longer DAST-10 screen was also 100% sensitive (95% CI, 90.6%-100%) for the detection of a current drug use disorder and was 77% specific (95% CI, 71.5%-81.9%); overall, its test characteristics were similar to those of the single-question screen (**Table 3**). Participant education and primary language affected point estimates of the sensitivity and specificity of the single-item screen very little, although for some groups with small sample sizes there was insufficient power to exclude large differences (**Table 4**). The single-item screen may be less specific for the detection of a current drug use disorder in men and in Hispanic patients.

COMMENT

A single-question screen was sensitive and specific for the detection of drug use and drug use disorders in a sample of primary care patients. Its test characteristics were simi-

Table 4. Single-Question Screen for the Detection of Current Drug Use, in Selected Subgroups

Characteristic	No.	Sensitivity, % (95% CI)	Specificity, % (95% CI)	AUC
Female	155	100 (61.0-100)	80.5 (73.4-86.1)	0.93
Male	131	100 (89.0-100)	63.0 (53.2-71.8)	0.92
Non-Hispanic, white	45	100 (74.1-100)	79.4 (63.2-89.6)	0.94
Non-Hispanic, black	176	100 (80.6-100)	73.8 (66.4-80.0)	0.92
Hispanic	46	100 (70.1-100)	59.5 (43.5-73.6)	0.91
English primary language	223	100 (89.3-100)	72.8 (66.1-78.6)	0.93
English not primary language	63	100 (56.6-100)	75.9 (63.5-85.0)	0.92
High school graduate	205	100 (85.1-100)	74.3 (67.5-80.1)	0.91
Not high school graduate	81	100 (79.6-100)	71.2 (59.4-80.7)	0.95

Abbreviations: AUC, area under the receiver operating curve; CI, confidence interval; current, within the past year (12 months); LR, likelihood ratio.

lar to those of a longer screening tool in this sample, as well as in other studies reported in the literature.[4]

Drug use is prevalent in primary care.[1] While national guidelines do not currently recommend universal screening for drug use in primary care, recent evidence supports the effectiveness of brief intervention in this setting, and screening, brief intervention, and referral to treatment initiatives are widespread.[3,17] In addition to identifying patients who might benefit from brief physician counseling, drug use screening is likely worthwhile in many clinical circumstances, such as identifying potential medication interactions and prescribing risks (as when clinicians ask patients to report prescription and over-the-counter medication use and alternative medicines as part of routine care).

Time constraints in the primary care setting have been cited as a reason for failure to provide screening and prevention in general (according to one estimate, providing all recommended preventive services to an average primary care panel would require 7.4 hours out of each workday[18]). Successful screening and brief intervention programs therefore require a means of quickly selecting, from among all primary care patients, those most likely to benefit from further assessment and intervention. Single-question screening tests for unhealthy alcohol use have been validated, and one such test is currently recommended by the National Institute on Alcohol Abuse and Alcoholism in its most recent clinician's guide.[6] To our knowledge, no other single-question screening test for drug use has been validated in any setting. Such a screening test could facilitate early identification and brief intervention, as well as the avoidance of prescription errors and associated risks.

A number of drug use screening instruments have been proposed for use in general medical settings, ranging from 2 questions to more than 70.[5,19] Some of these are modified versions of alcohol screening tests, and some ask simultaneously about both alcohol and drugs (so-called conjoint screens). Conjoint screens may be more acceptable to some patients than direct questioning about drug use but also require more clarification of a positive screen result, and some of the questions, adapted from alcohol screening tests, may be less applicable to drug use (eg, the "eye-opener" question from the CAGE-AID ["Adapted to Include Drugs"] questionnaire).[20] A brief, 2-item conjoint screen (TICS) has been validated, representing a screening strategy of equivalent brevity to asking a single question about drug use and a single question about alcohol. The TICS was 79% sensitive and 78% specific for either an alcohol or drug use disorder. The sensitivity for a drug use disorder was similar, but specificity was not reported.[19] Two longer, but still brief, conjoint screens, the CAGE-AID and RAFFT, have been tested in adults, with similar test characteristics.[20,21] These conjoint tests target drug disorders but do not specifically identify drug use.

The DAST-10 (to our knowledge, not validated in a primary care sample until the present article), the Drug Use Disorders Identification Test (validated only in criminal justice and detoxification settings), and the Alcohol, Smoking and Substance Involvement Screening Test (ASSIST), 3 screening questionnaires that ask about drug use specifically, have better test characteristics than the shorter conjoint screening tests and address part of the spectrum of clinical interest beyond drug diagnoses to include use and problems, but their length (10-28 questions for the DAST-10 and >70 questions for the ASSIST) and the need for scoring represent significant barriers to their use as screens in the primary care setting.[4,5,22] As a screening test (as opposed to an assessment of severity or a diagnostic tool), the single-question screen performed almost as well as the longer DAST-10 in the sample that we studied. Longer screening tools may, however, have promise as electronic medical record systems with decision support become more widespread (and as evidence for the validity of the ASSIST accumulates), potentially as a follow-up assessment after a positive single-question screen result, or even as a written previsit questionnaire. In summary, in terms of brevity, ease of scoring, and validity for detecting the spectrum of drug use conditions of interest in primary care, and therefore, likely greater widespread implementation, the single-question screen seems to have favorable characteristics.

For a screening test for drug use to be useful, it must be applicable to the broad range of people seen in primary care. The diversity of our participant sample allowed us to examine the effect of sex, ethnicity, primary language, and education on the accuracy of the single-question screen. While variations were seen in the sensitivity and specificity of the test across these groups, the differences were small.

Our study has several limitations. A higher than expected proportion of participants reported substance use disorders, likely reflecting the fact that they were re-

cruited from an urban safety-net hospital located in a community where the prevalence of such problems is high. While this potentially limits the generalizability of our results, it is this type of high-risk population that is typically targeted for screening and brief intervention (as mentioned, universal screening of all adults is not currently recommended, whereas targeted screening is recommended).[23] Nevertheless, additional study of the screening question in other settings (as well as in other language and in written and computer-based versions) is warranted. Participants were also assured anonymity, a condition that improves the accuracy of the reference standard interview but that may also serve to overestimate the accuracy of the screening test itself. This is consistent, however, with the methods of most other studies of screening tests for substance use disorders, thus allowing comparability of our findings with those of other studies.

The single-question screen accurately identified primary care patients who use drugs. Some patients who have positive tests results will have severe drug use disorders requiring referral to substance abuse treatment, while those who use drugs but have not experienced severe health or interpersonal problems might benefit from brief intervention by the primary care provider. The lack of an efficient way to distinguish these 2 groups is a challenge that must be addressed when implementing screening for drug use. The DAST-10 and the ASSIST screening tools, in providing scores, provide a measure of severity. Even though they may be too long for universal screening in many settings, they might be used for assessments after a single-item screening question is answered in the affirmative. However, this approach has not been tested or validated.

The single-question screen accurately identified a broad spectrum of drug use in this sample of primary care patients. The sensitivity and specificity of this single question was comparable with that reported for longer instruments in other studies. These findings support the use of this brief screen when identification of drug use is desired in primary care settings, which should, in turn, facilitate the implementation of screening and brief intervention programs in this setting.

Accepted for Publication: December 14, 2009.
Correspondence: Peter C. Smith, MD, MSc, Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, Second Floor, Crosstown Center, 715 Albany St, Boston, MA 02118 (peter.smith@bmc.org).
Author Contributions: *Study concept and design:* Smith and Saitz. *Acquisition of data:* Smith, Schmidt, and Saitz. *Analysis and interpretation of data:* Smith, Allensworth-Davies, and Saitz. *Drafting of the manuscript:* Smith, Schmidt, and Saitz. *Critical revision of the manuscript for important intellectual content:* Smith, Allensworth-Davies, and Saitz. *Statistical analysis:* Allensworth-Davies. *Obtained funding:* Smith and Saitz. *Administrative, technical, and material support:* Schmidt, Allensworth-Davies, and Saitz. *Study supervision:* Smith and Saitz.
Financial Disclosure: None reported.
Funding/Support: This research was supported by a grant from the National Institute on Alcohol Abuse and Alcoholism (R01-AA010870).
Previous Presentations: This study was presented in abstract form at the Addiction Health Services Research Conference; October 22, 2008; Boston, Massachusetts; and at the annual meeting of the Association for Medical Education and Research in Substance Abuse; November 8, 2008; Washington, DC.

REFERENCES

1. Cherpitel CJ, Ye Y. Drug use and problem drinking associated with primary care and emergency room utilization in the US general population: data from the 2005 national alcohol survey. *Drug Alcohol Depend.* 2008;97(3):226-230.
2. Saitz R, Mulvey KP, Plough A, Samet JH. Physician unawareness of serious substance abuse. *Am J Drug Alcohol Abuse.* 1997;23(3):343-354.
3. 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):280-295.
4. Yudko E, Lozhkina O, Fouts A. A comprehensive review of the psychometric properties of the Drug Abuse Screening Test. *J Subst Abuse Treat.* 2007;32(2):189-198.
5. WHO ASSIST Working Group. The Alcohol, Smoking and Substance Involvement Screening Test (ASSIST). *Addiction.* 2002;97(9):1183-1194.
6. National Institute on Alcohol Abuse and Alcoholism. *Helping Patients Who Drink Too Much: A Clinician's Guide, 2005 Edition.* Bethesda, MD: National Institute on Alcohol Abuse and Alcoholism; 2007.
7. 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):783-788.
8. Blanchard KA, Morgenstern J, Morgan TJ, Lobouvie EW, Bux DA. Assessing consequences of substance use: psychometric properties of the inventory of drug use consequences. *Psychol Addict Behav.* 2003;17(4):328-331.
9. Saitz R, Allensworth-Davies D, Cheng DM, Smith PC, Samet JH. Reliability and validity of the Short Inventory of Problems Modified for Drug Use [abstract 524]: 71st Annual Meeting of the College on Problems of Drug Dependence; June 29, 2009; Reno, Nevada. http://www.cpdd.vcu.edu/Pages/Meetings/CPDD09AbstractBook.pdf. Accessed April 3, 2010.
10. Kessler RC, Abelson J, Demler O, et al. Clinical calibration of DSM-IV diagnoses in the World Mental Health (WMH) version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI). *Int J Methods Psychiatr Res.* 2004;13(2):122-139.
11. Cone EJ, Presley L, Lehrer M, et al. Oral fluid testing for drugs of abuse. *J Anal Toxicol.* 2002;26(8):541-546.
12. Niedbala RS, Kardos K, Fries T, Cannon A, Davis A. Immunoassay for detection of cocaine/metabolites in oral fluids. *J Anal Toxicol.* 2001;25(1):62-68.
13. Niedbala RS, Kardos K, Waga J, et al. Laboratory analysis of remotely collected oral fluid specimens for opiates by immunoassay. *J Anal Toxicol.* 2001;25(5):310-315.
14. Niedbala RS, Kardos KW, Fritch DF, et al. Detection of marijuana use by oral fluid and urine analysis following single-dose administration of smoked and oral marijuana. *J Anal Toxicol.* 2001;25(5):289-303.
15. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology.* 1982;143(1):29-36.
16. Altman DG, Gardner MJ. Confidence intervals for research findings. *Br J Obstet Gynaecol.* 1992;99(2):90-91.
17. Humeniuk R, Dennington V, Ali R; World Health Organization ASSIST Phase III Study Group. *The Effectiveness of a Brief Intervention for Illicit Drugs Linked to the Alcohol, Smoking and Substance Involvement Screening Test (ASSIST) in Primary Health Care Settings.* Geneva, Switzerland: World Health Organization; 2008.
18. Yarnall KSH, Pollak KI, Ostbye T, Krause KM, Michener JL. Primary care: is there enough time for prevention? *Am J Public Health.* 2003;93(4):635-641.
19. Brown RL, Leonard T, Saunders LA, Papasouliotis O. A two-item conjoint screen for alcohol and other drug problems. *J Am Board Fam Pract.* 2001;14(2):95-106.
20. Hinkin CH, Castellon SA, Dickson-Fuhrman E, Daum G, Jaffe J, Jarvik L. Screening for drug and alcohol abuse among older adults using a modified version of the CAGE. *Am J Addict.* 2001;10(4):319-326.
21. Bastiaens L, Riccardi K, Sakhrani D. The RAFFT as a screening tool for adult substance use disorders. *Am J Drug Alcohol Abuse.* 2002;28(4):681-691.
22. Berman AH, Bergman H, Palmstierna T, Schlyter F. Evaluation of the Drug Use Disorders Identification Test (DUDIT) in criminal justice and detoxification settings and in a Swedish population sample. *Eur Addict Res.* 2005;11(1):22-31.
23. United States Preventive Services Task Force. *Screening for Illicit Drug Use*. Rockville, MD: Agency for Healthcare Research and Quality; 2008.

Screening and Brief Intervention for Drug Use in Primary Care

Leah E. Squires, MA, Daniel P. Alford, MD, MPH, Judith Bernstein, PhD, ADN, MSN, Tibor Palfai, PhD, and Richard Saitz, MD, MPH

Key Words: primary care, drug use, substance use, screening, brief intervention, motivational interviewing

(*J Addict Med* 2010;4: 131–136)

This clinical case involves a man with unhealthy alcohol and drug use who presents for an initial visit to primary care with complaints of heartburn and a recent admission for chest pain. Four expert clinicians contribute their thoughts about the case.

CASE DESCRIPTION

A 45-year-old man (Mr. M) hospitalized 3 months prior for chest pain ("noncardiac," myocardial infarction [MI] ruled out) presents to a primary care physician (PCP) with heartburn worse after meals and not helped by antacids. He denies weight loss, vomiting, or bloody stools/melena. He has some trouble falling asleep because of worries about his job and sometimes forgets to pick up his kids from their activities. He is married and has 2 children (a son aged 9 yr and a daughter aged 11 yr) and working as a manager at an electronics store. His last visit to a PCP was 10 years ago. His father had a heart attack at the age of 50 years. Three years ago, he was injured in a motor vehicle crash and had facial lacerations and rib fractures; urine toxicology was positive for cocaine and tetrahydrocannabinol. Physical examination is notable for blood pressure 152/94 mm Hg and an S4. An electrocardiogram is normal, and the rapid plasma reagin, thyroid stimulating hormone, vitamin B12 level, liver enzymes, renal function, and complete blood count are normal.

From the Department of Psychology (LES, TP), Boston University; Clinical Addiction Research and Education (CARE) Unit (DPA, RS), Section of General Internal Medicine, Boston Medical Center and Boston University School of Medicine; Department of Community Health Sciences (JB), Boston University School of Public Health; Department of Emergency Medicine (JB), Boston University School of Medicine; and Department of Epidemiology (RS), Boston University School of Public Health, Boston, Massachusetts.

Received for publication April 14, 2010; accepted August 3, 2010.

Send correspondence and reprint requests to Leah E. Squires, MA, Department of Psychology, Boston University, 648 Beacon Street, 4th Floor, Boston, MA 02215. e-mail: lsquires@bu.edu.

Supported in part by the National Institute on Drug Abuse (NIDA) (Grant no. 1R01DA025068) and NIDA Diversity Supplement (Grant no. 3RO1DA025068-02S1 to L.E.S.).

ISSN: 1921–0629/10/0403-0131
DOI: 10.1097/ADM.0b013e3181f59777

Screening

Primary care presents an opportunity to screen for lifestyle habits that may impair health and provide intervention when warranted. A majority of individuals with substance use problems do not recognize their use as problematic. Based on considerable empirical support for screening and brief intervention (SBI) for risky alcohol use, and evidence suggesting that SBI for other drugs may be beneficial (Babor and Kadden, 2005; Madras et al., 2009), Mr. M's primary care clinic uses the NIDA-modified Alcohol Smoking and Substance involvement Screening Test (nm-ASSIST) (http://www.drugabuse.gov/NIDAMED/screening/). The nm-ASSIST is an adaptation of the World Health Organization-developed ASSIST (World Health Organization ASSIST Working Group, 2002) designed for use in primary care settings. The nm-ASSIST is easily scored and provides recommendations for intervention based on drug-specific scores. It also includes screening information to identify those who may be hazardous/harmful (vs dependent) drinkers. In this clinic, patients complete the nm-ASSIST "prescreen," which asks about lifetime use of substances, as part of a self-report intake packet, which is reviewed by the office nurse. Positive endorsement of any prescreen item leads to completion of the rest of the nm-ASSIST as a nurse administered interview during the review of vital signs, allergies, pain symptoms, medications, and other preventive services that might be due.

On the prescreen, Mr. M indicates lifetime use of alcohol, tobacco, marijuana, and cocaine. During the interview, Mr. M states he has used marijuana daily and cocaine monthly in the past 3 months. He also reports heavy drinking episodes on the days when he uses cocaine. He reports quitting tobacco 5 years ago. On the nm-ASSIST, Mr. M's receives a cannabis use score of 17 and a cocaine use score of 6. These "moderate risk" (4–26) scores, along with Mr. M's episodic heavy drinking prompt recommendation for a brief intervention. Based on the results of screening, the nurse conducts a brief intervention with Mr. M before the doctor's examination.

Brief Intervention-Treatment Framework

As practiced in Mr. M's clinic, brief intervention consists of one 20- to 30-minute counseling contact with the office nurse and the option of a follow-up visit or phone call.

After the intervention, the nurse flags the patient's chart so that the PCP is aware that the brief intervention was conducted before seeing the patient. The flag consists of a simple checklist indicating (1) which substances were reviewed; (2) whether a change plan was completed; and (3) whether a follow-up visit was scheduled. The nurse applies principles of motivational interviewing (Miller and Rollnick, 2002) to establish a collaborative atmosphere in which the patient is encouraged to discuss his/her substance use in the context of this health visit.

Intervention

After the administration of the nm-ASSIST, the nurse asks Mr. M whether they can spend a few more minutes discussing his substance use and health. Mr. M is initially reluctant to engage in the conversation and states that he doesn't perceive his substance use as problematic. The nurse reassures Mr. M that the goal of the discussion is for his health care team to get a better sense of how the patient understands the role that substances play in his life and how substance use is related to his health. The nurse asks Mr. M to review his current health concerns. Mr. M states that he had difficulty sleeping and lies in bed worrying about his job security. Since his hospital admission 3 months ago (and his father's MI at a young age), he is also been concerned about his health and sometimes worries about how his family would cope if he did have a serious health condition. He notes that smoking marijuana provides stress relief. The nurse encourages Mr. M to describe other "benefits" of marijuana use. Mr. M, pleasantly surprised at being asked, states that marijuana helps him fall asleep, and that after he smokes, he feels able to put his worries aside for the evening. The nurse then asks Mr. M to discuss the positive aspects of using cocaine. Mr. M notes that he uses it only occasionally with his friends, snorting cocaine and drinking beers while watching sports or playing video games.

After discussing these benefits, the nurse asks Mr. M to describe the "downsides to using." Mr. M states that in the past few months he has experienced some conflict with his wife over his smoking. The nurse asks for elaboration on these events, including emotional consequences of this conflict. Mr. M also states that he doesn't want his kids to pick up his "bad habit." He also notes that since having a car accident 3 years ago, he no longer drives after spending time drinking and using cocaine with his friends. In addition, Mr. M mentions that he occasionally noticed his heart "racing" after using cocaine, but it doesn't bother him.

The nurse summarizes the pros and cons of marijuana and cocaine use described by Mr. M and then asks permission to share information about the possible impact of Mr. M's substance use on his health. The nurse is careful to present the feedback as potentially useful information rather than a prescription for change. The nurse also points out that Mr. M's current episodic heavy alcohol use also places him at increased risk for accidents, such as his past car accident. The nurse highlights associations between substance use and reported health concerns identified by Mr. M, specifically concerns about his memory and cardiac health. During the delivery of this information, the nurse engages the patient and asks for reactions, comments, and thoughts about what is being presented.

When asked what led to him changing his past behavior, Mr. M endorses finances, concern for his children, and his wife's concern as the main reasons he cut back on both drinking and other drug use. He says change was "not too hard" at the time because he was busy working and helping his wife take care of their first child in the evening. After summarizing Mr. M's past change efforts, the nurse asks Mr. M for his thoughts about his current use. Mr. M states that he'd like to cut back on his marijuana but doesn't really see the need to change his cocaine use because it is infrequent. He states that he was comfortable with his current alcohol consumption. Mr. M and the nurse work on developing a change plan. Mr. M decides to reduce his marijuana smoking by engaging in other stress relieving activities, including spending more time with his children and a friend who doesn't use drugs. He identifies his friend Charlie and his wife as people he can enlist to help with this plan. The nurse expresses appreciation to Mr. M for his involvement in the discussion and offers him a follow-up visit if he thinks it would be helpful to him. Mr. M accepts the offer of a follow-up visit in 2 weeks to review his progress. The nurse also provides Mr. M with brochures summarizing the health information they discussed (impact of smoking, cocaine, marijuana, and alcohol use on the body) and a copy of his written plan for change. The nurse informs Mr. M that his PCP will see their discussion noted in his medical record and encourages Mr. M to speak to the doctor if he has additional questions.

Follow-Up Visit

Mr. M arrives on time for his follow-up visit and reports that he hasn't used cocaine for the past couple of weekends. He's also cut back on his marijuana use to 1 to 2 evenings. However, his drinking increased on weekend evenings. Mr. M states that thinking about smoking marijuana as similar to smoking cigarettes and the possibility of consequences that might affect his children made cutting back "something I want to do;" however, he notes that he still struggles to fall asleep at night and has on occasion gone outside to smoke to fall asleep.

The nurse and Mr. M review his progress and reformulate his change plan. Mr. M restates his intent to reduce his marijuana use. At the end of the follow-up visit, the nurse provides Mr. M with information about brief substance use treatment offered through the hospital and encourages him to contact the primary care team if he needs additional help making changes.

Record Keeping

In addition to the flag created for the doctor, the nurse notes in Mr. M's medical record the results of the substance use screening, his intention to cut back on his own, his change plan, and perceived barriers and supports to change. As part of standard practice, Mr. M will be asked about his progress at his next primary care visit and will be rescreened on an annual basis.

DISCUSSION

Daniel P. Alford, MD, MPH, FACP, FASAM

This case highlights the multiple important opportunities to screen patients for unhealthy substance use in general healthcare settings. Emergency room visits and hospitalizations are likely "reachable moments" (Shanahan et al., 2010) for patients with unhealthy substance use, especially when there is a link between their acute illness and their substance use. Before his primary care visit, Mr. M had previous emergency rooms visits and a hospitalization. These were all opportunities for healthcare providers to identify and address Mr. M's unhealthy substance use and draw connections between his substance use and his chief complaints (ie, motor vehicle crash and chest pain). Despite unhealthy substance use being a common problem in general healthcare settings, physicians are often reluctant or are inadequately trained to effectively screen or provide brief interventions (Isaacson et al., 2000).

An important question when implementing substance use SBI in general healthcare settings is who should perform it and when should it take place? During Mr. M's visit to primary care, a nurse administered the SBI for those patients who "prescreened" positive. Collaborative care in which staff other than the PCP screen and counsel patients is certainly one way to lessen the burden of multiple preventive agendas for PCPs. However, in this case, I am concerned about the low PCP involvement. In this SBI, the PCP was informed of the screening results, brief intervention conducted, and the patient's plan for change before the PCP visit. Ideally, this would be used by the PCP to emphasize the association between the patient's substance use and personally relevant health outcomes, reinforcing the feedback that the patient received and supporting any change efforts that the patient may wish to undertake; however, it is not clear that this occurred. The timing is also an issue in implementing SBI in primary care settings. With valid single-item screening tests for alcohol (Smith et al., 2009) and drugs (Smith et al., 2010) now available, it is feasible to "prescreen" patients before they see their PCP. However, for patients prescreening positive, a more extensive screening/assessment (eg, NIDA-modified ASSIST) is needed, followed by a brief intervention, all of which can be time consuming. The challenge with the nurse model in this case is how readily it could be implemented in a busy primary care practice where 20 to 30 minutes before a PCP visit is not available and may interfere with patient flow. Delaying the full screen/assessment and brief intervention until after the PCP visit is one possible solution but risks the patient leaving without having their unhealthy substance use addressed and excludes the PCP from the SBI process.

Finally, when a patient such as Mr. M screens positive for multiple substances (ie, cocaine, marijuana, and alcohol), which substance should the brief intervention focus on? Should it focus on all the substances, the substance the patient is most concerned about, or the substance that is most risky for the patient's health as determined by the healthcare provider? In Mr. M's case, despite having a lower nm-ASSIST score for cocaine, I could argue that his "occasional" intranasal cocaine use is putting his health at grave risk. He likely has preexisting heart disease with his elevated blood pressure and fourth heart sound. Cocaine use can cause increased myocardial oxygen demand, vasoconstriction, thrombosis, and premature atherosclerosis, which in combination with his family history of coronary heart disease put him a high risk for having a MI (Lange and Hillis, 2001). The occurrence of MI with cocaine is unrelated to the amount ingested, route of administration, or frequency of use (Lange, 2003) Moreover, the combination of cocaine and alcohol produces cocaethylene (Laizure et al., 2003), which increases the risk of cocaine-associated cardiac toxicity. I wonder whether Mr. M's comment that he "doesn't see the need to change his cocaine use" because it was "infrequent" was based on not being adequately informed about the potential risks that cocaine poses for him. It would have been helpful to first assess Mr. M's understanding of the potential health risks of his cocaine use before offering him feedback and information. Although it is important for brief interventions to be patient centered allowing patients to set goals, it would be unfortunate if the patient chose priorities for change without being fully informed of the risks associated with his/her substance use.

Judith Bernstein, PhD, ADN, MSN

Sometime it takes centuries to discover the obvious. As long ago as 1670, the French mathematician Pascal said, "people are generally better persuaded by the reasons which they have themselves discovered than by those which have come into the mind of others." Clinicians spend much valuable time lecturing patients about the negative imperatives for behavior change—risk factors and potential health consequences—but often overlook the uses of patient self-reflection and positive reinforcement. Motivational interviewing provides a toolbox for this type of patient-centered brief intervention in the course of a health care visit. Screening for unhealthy alcohol use is currently recommended in the primary care setting by the U.S. Preventive Services Task Force (2004), and moderate effect sizes for reductions in both quantity and alcohol-related consequences are well substantiated by a series of meta-analyses (Kaner et al., 2009). The evidence is not yet in for drug use, but preliminary research and program evaluation findings suggest promise (Bernstein et al., 2005; Madras et al., 2009).

As this case illustrates, motivational interviewing reflects "the rather simple notion that the way clients are spoken to about changing addictive behavior affects their willingness to talk freely about why and how they might change" (Rollnick, 2001). The nurse began by asking permission to discuss the subject, establishing respect for the patient and setting the stage for the patient to be the engine of change. The nurse then elicited the patient's perspective on the pros and cons of his substance use and only then offered nonjudgmental feedback about the possibility of connection between his substance use and his health concerns. If the nurse had begun by asking Mr. M what he understood about the risks of use, resistance and denial might have been encountered. Instead, by starting by eliciting positive effects of drug and alcohol use, the nurse was able to learn about the role cocaine

and alcohol played in important relationships and the patient's use of marijuana for stress reduction. Mr. M was then able to weigh these "benefits" against risks that he named for himself; the result was a plan for change that he believed he had the capacity to carry out. When he reported at follow-up that he had accomplished some but not all of these goals, he was received with positive reinforcement, rather than embarrassed about having "slipped." Brief intervention of this type is not a substitute for specialized treatment in those who exhibit dependence, but it can lead to behavior change over time, especially when reinforced in the context of continued primary care. Mr. M may still need a referral for counseling to examine sleep disruption and stress processing, but he has already changed his thinking about drug use and begun to alter negative health behaviors.

It is important to note that a nurse delivered this intervention. PCPs are often reluctant to begin complex conversations about behavior change because of time pressures. Use of a physician extender, a nurse, social worker, or an outreach worker who functions as a member of the clinical team could ensure that fulfilling U.S. Preventive Services Task Force (2004) recommendations does not take the 7.4 hours of physician clinical time each day that a landmark time-motion study demonstrated (Yarnall et al., 2003). Identification of unhealthy substance use during a clinical visit, followed by intervention to encourage healthier behaviors, is too important an opportunity to miss.

Tibor Palfai, PhD

With empirical support for the use of alcohol SBI in primary care (Kaner et al., 2009), investigators have adopted similar approaches for drug use with promising results (Madras et al., 2009). Although there are a variety of brief interventions that may be used to address substance use in medical settings (Babor and Kadden, 2005; Kaner et al., 2009), motivational interviewing (Miller and Rollnick, 2002) is particularly well suited for addressing substance use in the primary care setting where patients often do not identify their substance use as problematic or perceive the need to change. Moreover, motivational interviewing and its adaptations have received empirical support as a brief treatment for illicit drug use and marijuana use specifically (Hettema et al., 2005).

In this case, the patient presents with hazardous alcohol use, cocaine use, and frequent marijuana use. Among the challenges faced by the nurse is how to enhance motivation to change marijuana use in a patient who appears to view his use as nonproblematic and how to do this in a manner that may be enduring to support behavior change (eg, self-change and linkage to treatment). One of the features of motivational interviewing that makes it particularly valuable for this case is its emphasis on supporting patient autonomy (Markland et al., 2005). There is now considerable empirical support for the view that behavior change is facilitated to the extent that patients perceive their activity as autonomously controlled (Ryan and Deci, 2008). Autonomous sources of behavior include personally important values, goals, and beliefs and reflect the degree to which one experiences an action as willingly chosen. Autonomy support occurs through a number of processes including understanding and acknowledging perspectives, expressing regard, minimizing pressure and control, providing choices, and delivering a meaningful rationale for suggestions or requests.

The interview illustrates the variety of ways that the nurse provides autonomy support for the patient through the intervention. As detailed above, the nurse skillfully addresses initial patient reluctance by presenting the interview as a collaborative discussion that is intended to better understand substance use from the patient's point of view. By inviting the patient to first describe his current health concerns, the nurse is able to acknowledge what is most important for the patient, express empathy, and help the patient consider substance use in the context of overall health. The nurse's efforts to support autonomy are evident in the way that feedback about use and information about potential consequences are presented to the patient. These aspects of the intervention, which are common to a number of adaptations of motivational interviewing (Burke et al., 2003), may be particularly relevant for patients who are not aware of the association between substance use and presenting health concerns. Appreciating that the delivery of information about negative health effects may be perceived by the patient as an attempt to convince him to change, the nurse asks permission to deliver information about the effects of marijuana and cocaine on health, states that the information may or may not be experienced as important for the patient, and frequently checks with the patient about his perception of the information. Statements that indicate that the nurse values his perspective and appreciates his engagement are particularly important during this part of the intervention.

The nurse also seeks to identify, elaborate on, and reinforce instances in which the patient has been successful in promoting change. This includes both efforts to identify patient successes to enhance self-efficacy regarding current change, as well as efforts to highlight and elaborate on goals and values that have served as important sources of change in the past. As this patient indicates an interest in change, the nurse discusses options about how he might undertake change efforts and elicits sources of support and barriers to change. In addition, the nurse offers a follow-up session, which is presented as an option that the patient may use if helpful to him. The nurse further emphasizes choice and autonomy in the manner in which a follow-up session is offered in person or by phone. This follow-up session allows the nurse to explore the patient's experience with behavior change and continue support for autonomy while keeping the patient engaged with the change goal in a manner that he has chosen.

Richard Saitz, MD, MPH, FACP, FASM

It is noteworthy that Mr. M's PCP has implemented SBI for drugs. If drug SBI can improve primary care patient outcomes, then she/he will have been at the vanguard of practice. However, all universal preventive practices have opportunity costs, so each new practice should have proven efficacy. Guessing about efficacy, even when the practice seems logical, is not good enough—many good ideas have been later proven ineffective. We know that SBI for nondependent unhealthy alcohol use has modest efficacy (U.S. Preventive Services Task Force 2004); data are lacking for

single primary care visit drug SBI (Saitz et al., 2010). Also, there are reasons why it may not work, such as severity of those who use certain drugs, patient preferences against being referred to treatment, the wide range of substances used, and the complexity of prescription drug abuse. Brief adaptations of motivational interviewing similar to those used in this case are most likely to work.

In Mr. M's case, the practice wasn't screening (or its synonym "prescreening," whieh has somehow entered the vernacular in SBI programs but has no meaning distinct from "screening"). Screening quickly separates those at risk from those at lower risk and implies that the patient is asymptomatic with regard to the target condition. Yet, Mr. M had a hospitalization for chest pain, multiple trauma and positive urine drug tests, has insomnia and heartburn, and is sometimes forgetful. Screening was complete (and positive) before any questions were asked during this visit. Appropriate evaluation for Mr. M would include assessment relevant to the differential diagnosis of his symptoms and conditions, including alcohol and other drug use, which can cause or play a role in each.

Fortunately, Mr. M's PCP has implemented the nm-ASSIST, despite its length, need for scoring, less immediately relevant (to screening) questions about lifetime use, and inability to specify whether a patient is drinking risky amounts. Why then, fortunate? Because the benefit of such a tool is the assessment data it provides. Answers to individual questions give the clinician something to discuss. Scores give an estimate of severity and thus can inform brief intervention goals and urgency. Another fortunate feature of Mr. M's PCP's practice is having a nurse who can do brief counseling. This uncommon setup may be ideal for addressing a range of health behaviors in primary care. If drug SBI is proven efficacious, staff with behavioral expertise could step in after brief (eg, single question) validated screening. This approach seems most feasible and could result in wider dissemination of SBI.

I was surprised at Mr. M's response to the brief intervention. Although Babor (2004) found efficacy for marijuana brief intervention, their study was of people who sought the intervention. However, screening-identified primary care patients often do not perceive consequences or risks and have little ambivalence about their marijuana use for clinicians to work with. I might have prioritized differently. Mr. M had heartburn and hypertension, which could have been related to alcohol consumption, and chest pain (and possibly premature atherosclerosis) that could have been related to his cocaine use (and heart racing, which was related to it). I might have emphasized the link between these conditions of importance to him, and his alcohol and cocaine use, highlighting the discrepancy between what he values and his use, rather than spend precious time on his marijuana use.

The ability to make links between substanee use and medical conditions raises the question of the physician role in SBI. The best evidence for efficacy of alcohol SBI is for a PCP eounseling his or her patient on multiple occasions (Whitlock et al., 2004). If the physician isn't going to do the screening or the brief intervention, they should at least provide feedback and a warm handoff.

Finally, this case hints at another challenge for PCPs and patients. Information in medical records can have impact on insurability and potentially employment if the patient releases their records, which they often must do when applying for benefits. Federal substance dependence confidentiality regulations apply to programs that hold themselves out as providing treatment for addiction. In this case, there is no diagnosis of addiction, and the PCP was not representing the practice as an addiction treatment provider, so the regulations don't apply. The good news is that this PCP will be able to provide a safe, comprehensive eare for the patient including his substance use and other medical conditions. The bad news is that there will be potentially damaging information in the record. Privacy protections in primary care and addiction treatment programs should be reconsidered to prevent discrimination while encouraging high quality integrated healthcare.

REFERENCES

Babor TF. Brief treatments for cannabis dependence: findings from a randomized multisite trial. *J Consult Clin Psychol* 2004;72:455–466.

Babor TF, Kadden RM. Screening and interventions for alcohol and drug problems in medieal settings: what works? *J Trauma* 2005;59:S80–S87.

Bernstein J, Bernstein E, Tassiopoulos K, et al. Brief motivational intervention at a clinie visit reduces eocaine and heroin use. *Drug Alcohol Depend* 2005;77:49–59.

Burke BL, Arkowitz H, Menchola M. The efficacy of motivational interviewing: a meta-analysis of controlled clinieal trials. *J Consult Clin Psychol* 2003;71:843–861.

Hettema J, Steele J, Miller WR. Motivational interviewing. *Annu Rev Clin Psychol* 2005;1:91–111.

Isaacson JH, Fleming M, Kraus M, et al. A national survey of training in substance use disorders in residency programs. *J Stud Alcohol* 2000; 61:912–915.

Kaner EF, Dickinson HO, Beyer FR, et al. The effectiveness of brief alcohol interventions in primary carc settings: a systematic review. *Drug Alcohol Rev* 2009;28:301–323.

Laizure SC, Mandreli T, Gades NM, et al. Cocaethylene metabolism an interaction with cocaine and ethanol: role of carboxylcsterases. *Drug Metab Dispos* 2003;31:16–20.

Lange RA. Cocaine and myocardial infarctions. *Adv Stud Med* 2003;3:448–454.

Lange RA, Hillis LD. Cardiovascular complications of cocaine use. *N Engl J Med* 2001;345:351–358.

Madras BK, Compton WM, Avula D, et al. Screening, brief interventions and 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;99:280–295.

Markland D, Ryan RM, Tobin VJ, et al. Motivational interviewing and self-determination theory. *J Soc Clin Psychol* 2005;24:811–831.

Miller WR, Rollnick S. Motivational Interviewing: Preparing people for change. 2nd ed. New York, NY: Guildford Press, 2002.

Rollnick S. Principles of motivational interviewing. *Addiction* 2001;96: 1769–1770.

Ryan RM, Deci EL. A self-determination theory approach to psychotherapy: the motivational basis for effective change. *Can Psychol* 2008;49:186–193.

Saitz R, Alford DP, Bernstein J, et al. Screening and brief intervention for unhealthy drug use in primary care settings: randomized elinical trials are needed. *J Addict Med* 2010;4:123–130.

Shanahan CW, Beers D, Alford DP, et al. A transitional opioid program to engage hospitalized drug users. *J Gen Intern Med* 2010;25:803–808.

Smith PC, Schmidt SM, Allensworth-Davies D, et al. Primary care validation of a single-question alcohol screening test. *J Gen Intern Med* 2009;24:783–788.

Smith PC, Schmidt SM, Allensworth-Davies D, et al. A single-question

screening test for drug use in primary care. *Arch Intern Med* 2010;170:1155–1160.

U.S. Preventive Services Task Force. Screening and behavioral counseling interventions in primary care to reduce alcohol misuse: recommendation statement. *Ann Intern Med* 2004;140:554–556.

Whitlock EP, Polen MR, Green CA, et al. Behavioral counseling interventions in primary care to reduce risky/harmful alcohol use by adults: A summary of the evidence for the U.S. Preventive Services Task Force. *Ann Intern Med* 2004;140:557–558.

World Health Organization ASSIST Working Group. The Alcohol, Smoking, and Substance Involvement Screening Test (ASSIST): development, reliability and feasibility. *Addiction* 2002;97:1183–1194.

Yamall KS, Pollack KI, Ostbye T, et al. Primary care: is there enough time for prevention? *Am J Public Health* 2003;93:635–641.

Systematic Review: Treatment Agreements and Urine Drug Testing to Reduce Opioid Misuse in Patients With Chronic Pain

Joanna L. Starrels, MD, MS; William C. Becker, MD; Daniel P. Alford, MD, MPH; Alok Kapoor, MD, MSc; Arthur Robinson Williams, MA; and Barbara J. Turner, MD, MSEd

Background: Experts recommend opioid treatment agreements and urine drug testing to reduce opioid analgesia misuse, but evidence of their effectiveness has not been systematically reviewed.

Purpose: To synthesize studies of the association of treatment agreements and urine drug testing with opioid misuse outcomes in outpatients with chronic noncancer pain.

Data Sources: MEDLINE, PsycINFO, EMBASE, Cochrane Central Register of Controlled Clinical Trials (January 1966 to June 2009), reference lists, and expert contacts.

Study Selection: Original research addressing opioid medications, chronic pain, and treatment agreements or urine drug testing, with a sample size of 50 participants or more and published in English, Spanish, or French.

Data Extraction: Two investigators independently identified eligible studies, extracted data, and assessed study quality. The outcome of opioid misuse was defined as drug abuse, drug misuse, aberrant drug-related behavior, diversion, or addiction.

Data Synthesis: Of 102 eligible studies, 11 met inclusion criteria; 6 were in pain clinics and 5 were in primary care settings. Four primary care studies examined multicomponent strategies that included interdisciplinary support. All studies were observational and rated as poor to fair quality. In 4 studies with comparison groups, opioid misuse was modestly reduced (7% to 23%) after treatment agreements with or without urine drug testing. In the other 7 studies, the proportion of patients with opioid misuse after treatment agreements, urine drug testing, or both varied widely (3% to 43%).

Limitations: Diversity of interventions and opioid misuse measures precluded meta-analysis. Most studies evaluated combinations of interventions.

Conclusion: Relatively weak evidence supports the effectiveness of opioid treatment agreements and urine drug testing in reducing opioid misuse by patients with chronic pain. Further research on effective ways to monitor and reduce opioid misuse is needed, especially in primary care settings.

Primary Funding Source: Substance Abuse and Mental Health Services Administration, National Institute on Drug Abuse, and Robert Wood Johnson Foundation.

Ann Intern Med. 2010;152:712-720. www.annals.org
For author affiliations, see end of text.

Chronic pain is one of the most common reasons for medical visits (1, 2), affecting 20% to 50% of patients who visit primary care providers (3–5). During the past 2 decades, opioid analgesics have been increasingly prescribed for chronic noncancer pain (1, 6, 7) and are now among the most frequently dispensed medications in the United States (8). The expansion of opioid prescriptions for chronic pain was initially supported by research from the 1980s and early 1990s, which reported a low risk for opioid addiction (9). However, treatment of chronic pain with opioids has continued to increase, despite a lack of rigorous research demonstrating the effectiveness of long-term treatment (10–13) and a burgeoning public health threat posed by opioid misuse, including abuse, addiction, diversion, and unintentional overdose (14–18).

See also:

Print
Editors' Notes 713
Editorial comment. 747

Web-Only
Appendix Tables
Conversion of graphics into slides

To mitigate the risks for prescription opioid misuse, medical societies and regulatory agencies have published consensus guidelines recommending specific risk-reduction strategies, including written opioid treatment agreements and urine drug testing for patients with chronic pain who are prescribed long-term opioid analgesics (19–21). Because the adverse consequences of prescription opioid misuses including overdose or death, are severe and the risk for misuse cannot be reliably predicted, experts advise that "universal precautions" should be adopted when prescribing long-term opioid analgesics (22). Although most patients with chronic pain are managed in primary care practices (1), adoption of opioid treatment agreements and urine drug testing by primary care physicians has been limited. In 3 primary care studies of management of patients who were prescribed long-term opioids, only 23% to 44% of physicians completed treatment agreements with these patients, and only 8% to 30% obtained urine drug tests (23–25). Among the several reasons for limited use of these approaches by primary care physicians may be the lack of a clear evidence base for their effectiveness in reducing opioid misuse and other adverse outcomes (26, 27). This systematic review addresses this gap by evaluating the effectiveness of opioid treatment agreements and urine drug testing in reducing opioid misuse among outpatients with chronic noncancer pain.

Methods

Data Sources and Searches

We consulted with a research librarian and conducted a systematic literature search of MEDLINE, PsycINFO, EMBASE, and the Cochrane Central Register of Controlled Trials from the earliest index date to week 2 of June 2009 for relevant studies published in English, Spanish, or French. We used Medical Subject Headings and keywords to search opioid medications, chronic pain, and opioid treatment agreements or urine drug testing (**Appendix Table 1**, available at www.annals.org), combined by using the Boolean operator "and," as previously described (10, 28). To identify additional studies, we reviewed all reference lists from included studies, consensus statements, and relevant review articles. In addition, we consulted authors of included studies and experts in pain and opioid management.

Study Selection

Two reviewers screened abstracts identified by the literature search for potential eligibility. Articles were considered eligible for full review if they addressed opioid misuse outcomes after implementation of treatment agreements or urine drug testing; reported data from an original study (not a review or an editorial); were conducted in an outpatient setting in patients with chronic noncancer pain; included 50 or more participants; and were published as full manuscripts in English, Spanish, or French. Studies were not excluded on the basis of study design alone. We excluded studies in emergency departments because they do not provide longitudinal care, as well as studies of patients with cancer-related, acute, or postoperative pain. In a random sample of 100 abstracts, the observed agreement between reviewers on this first screening was 95%, which corresponded to substantial agreement ($\kappa = 0.76$).

To define the final set of eligible studies, 2 physician–researchers retrieved and independently reviewed full manuscripts. Eligible studies were conducted in a relevant study population and included patients who were prescribed at least 3 months of opioids or who received opioid treatment described as "long-term" or "chronic." Studies of opioid treatment agreements had to describe a written document that used any of 3 terms for the medication (opioid, controlled substance, or pain medication) and any of 4 terms for the type of document (agreement, contract, treatment plan, or informed consent). Because of the wide variation in the description and inclusion of the treatment agreement documents in identified studies, we did not further specify the type of treatment agreements for inclusion in the review. Studies of urine drug testing needed to use testing routinely as part of a management strategy and not as a single test to detect misuse for research purposes. The urine drug test needed to assess for controlled substances, but we did not define the type of assay or the specific drugs tested.

Context

Some experts recommend treatment agreements and urine drug testing to discourage misuse when opioids are prescribed for chronic noncancer pain, even though the effectiveness of these strategies is uncertain.

Contribution

This systematic review identified 11 poor- to fair-quality observational studies of treatment contracts and urine drug testing and found only weak, heterogeneous evidence that these strategies were associated with less misuse.

Implication

Careful studies of strategies to promote safe use of opioids for chronic noncancer pain are needed. Unless evidence becomes available to support a benefit of treatment contracts and urine testing, the absence of these strategies should not be considered a mark of poor-quality care.

—The Editors

With regard to opioid misuse outcomes, eligible studies needed to evaluate behaviors described as aberrant or indicative of abuse, misuse, or diversion, consistent with the terminology recommended by Ballantyne and LaForge (29). Examples of such behaviors include overuse of prescribed opioids, repeatedly lost or stolen prescriptions, receipt of opioid medications from multiple providers or sources, prescription adulteration or forgery, or concomitant use of nonprescribed controlled substances (licit or illicit) (29). Outcomes could have been measured from patients, providers, medical charts, or laboratory tests. Urine drug tests indicative of opioid misuse included those that were positive for a nonprescribed controlled substance or negative for the prescribed controlled substance and confirmed with gas or liquid chromatography and mass spectrometry. The observed agreement between reviewers was 94.2% ($\kappa = 0.83$). A third reviewer resolved disagreements about eligibility.

Data Extraction and Quality Assessment

Two reviewers used a standardized instrument to independently extract data from each study. They extracted data on the patient sample, outpatient setting, intervention, opioid misuse outcome measures, and results.

By using a quality assessment checklist that was derived from the ECRI Institute (13), Downs and Black (30), and Jadad and coworkers (31), 2 reviewers assessed the quality of included studies in 3 domains: sample, design, and a global assessment (32) (**Table 1**). We modified these measures to address features of observational studies. The quality checklist assigned a maximum of 15 points to each study: 3 for sample (external validity) and 12 for design (internal validity). In addition, each study was evaluated by using the modified GRADE (Grading of Recommendations Assessment, Development, and Evaluation) quality

Table 1. **Assessment of Study Quality**

Sample domain (maximum 3 points)*
- Persons with diverse types of chronic pain conditions included
- Persons with history of substance abuse included
- Persons with mental health disorders included

Design domain (maximum 12 points)*
- Prospective design
- Control group included
- Control participants from a similar population
- Intervention described clearly
- Intervention consistent among groups
- Outcome described clearly
- Outcome objective
- Completion or response rate ≥85%
- Distribution of potential confounders provided
- Multivariate analysis conducted
- Adequate adjustment for confounding
- Results clearly presented

Global assessment (maximum 4 points)†
- Very low = 1
- Low = 2
- Moderate = 3
- High = 4

* 1 point was assigned for each.
† Based on the GRADE (Grading of Recommendations Assessment, Development, and Evaluation) algorithm (33).

assessment criteria, which consider the rigor of the design, biases, and limitations on a scale of 1 to 4 (33). On the basis of scores on the quality checklist and GRADE, we assigned 1 of the following ratings to each study: excellent (score ≥11 and GRADE ≥3), good (score ≥11 and GRADE ≥2), fair (score = 6 to 10 and GRADE ≥2), or poor (score ≤5 or GRADE = 1). Ratings between reviewers had 73% agreement ($\kappa = 0.48$), and a third independent reviewer resolved disagreements in quality assessment. We did not exclude studies on the basis of quality. **Appendix Tables 2** and **3** (available at www.annals.org) provide a detailed quality assessment for each study, including scores for individual domains.

Data Synthesis and Analysis

For each study, we calculated the absolute risk reduction (ARR) and 95% CIs for opioid misuse in the intervention versus control or comparison group. Consistent with options described in the Cochrane Handbook for Systematic Reviews (34), we dichotomized a continuous outcome variable in 1 study to calculate the ARR; the study reported the mean number of opioid prescribers per patient per quarter, and we selected a cutoff of 2 or more as indicative of receiving opioids from multiple sources (35). The interventions evaluated and the measures of opioid misuse were too variable to permit meta-analysis, so we summarized results descriptively.

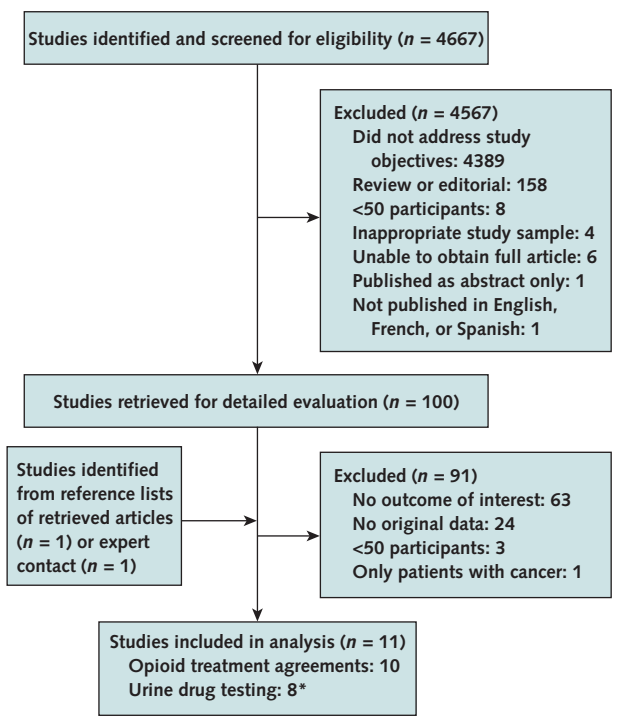
Role of the Funding Source

The Substance Abuse and Mental Health Services Administration, the National Institute on Drug Abuse, and the Robert Wood Johnson Foundation funded the study through the Treatment Research Institute's Program of Research to Integrate Substance Use into Mainstream Healthcare (PRISM). The funding sources had no influence on the analyses, preparation of the manuscript, or the decision to submit the manuscript for publication.

RESULTS

Of 4667 abstracts identified, 100 studies met criteria for full review, and 2 additional articles were identified from reference lists and expert contacts. Of these 102 studies, 11 met our inclusion criteria (35–45) (**Figure**). No studies evaluated opioid overdose, abuse, or dependence, as they were defined by the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition*. Six studies were conducted in pain specialty settings (37, 38, 42–45). Of the 5 studies in primary care settings, 3 evaluated interdisciplinary programs that integrated care by pain specialists, psychiatrists, nurses, or pharmacists on a referral basis for patients suspected of opioid misuse (36, 39, 40). Another primary care study included a secretary-based patient tracking system (35). Overall, 7 studies evaluated opioid treatment agreements as part of multicomponent management strategies that included routine urine drug testing (36–41, 44), and 1 study evaluated urine drug testing alone (45). Disparate outcome measures of opioid misuse were used, as described in **Table 2**.

Figure. **Study flow diagram.**



* Seven studies evaluated management strategies incorporating both treatment agreements and urine drug testing.

The description of treatment agreements and urine drug testing interventions varied widely among studies. Six of 10 studies evaluating treatment agreements provided a copy of the agreement (35, 39–42, 44), but the remaining 4 studies offered little or no description of it. The same treatment agreement was used in 2 studies in the same setting (39, 40), yielding 5 agreements available for review. All 5 of these agreements listed the conditions necessary for continuation of opioid medications. Although these conditions varied, all agreements required patients to agree not to sell or share opioid medications and to provide urine or blood specimens for drug testing at their providers' request. In most agreements, patients agreed not to abuse illicit drugs or alcohol, obtain opioids from more than 1 provider or pharmacy, or request a refill before the previous prescription should have been completed. Although 3 agreements informed patients that opioid medications may not improve their pain or function, only 1 stated that a lack of response would indicate that these medications should be discontinued (42). Among the 8 studies that evaluated urine drug testing, the frequency of and indications for testing varied. Testing was described as "regular" or "random" in 4 studies (37–39, 41), and 1 study defined a protocol for testing before enrollment and on a monthly basis thereafter (40). Three studies described a minimum frequency of urine testing (at enrollment, annually, or both), with more frequent testing according to the physician's judgment (36, 44, 45). The assay used in urine drug testing interventions was specified in 5 of 8 studies; 1 study used an immunoassay screen alone (38), and 4 used confirmatory gas chromatography and mass spectrometry (39, 40, 44, 45).

All 4 studies that included a control group reported a decrease in opioid misuse after treatment agreements as part of the opioid management strategy (35–38). Two of these studies were conducted in Veterans Administration primary care clinics that provided monthly visits with a clinical pharmacist (36) or tracking by a secretary (35) as part of the monitoring program. One Veterans Administration study (36) found that opioid misuse among the 335 study participants decreased from 51.0% before the intervention to 28.1% afterward (ARR, 22.9% [95% CI, 17.3% to 28.7%]) (36). The other Veterans Administration study (35) reported a decrease from 47.3% to 26.4% (ARR, 20.9% [CI, 10.6% to 31.2%]) in opioid misuse among 209 intervention patients compared with no change in matched control participants. The remaining 2 controlled studies were conducted in the same sample of patients in a pain specialty setting and used historical control participants (37, 38). These studies found statistically significant decreases in obtaining opioid medications from an outside source from 17.8% to 9.2% (ARR, 8.6% [CI, 4.4% to 12.8%]) and in detection of illicit drug use from 22.5% to 16.0% (ARR, 6.5% [CI, 1.3% to 11.7%]). The quality of the 4 controlled studies was rated fair for 2 (35, 36) and poor for 2 (37, 38). Quality concerns include unrepresentative and high-risk samples (35–38), use of pre–post design (35, 36), and historical control groups (37, 38). **Appendix Tables 2** and **3** provide a detailed quality assessment.

Seven uncontrolled studies found wide variation in the proportion of patients with opioid misuse after treatment agreements, urine drug testing, or both (3% to 43%) (39–45). Of the 3 primary care studies, 2 were conducted in the same interdisciplinary referral-based program for patients with suspected substance abuse (39, 40) and reported that 32% of patients misused opioid medications, despite completing opioid treatment agreements. Another uncontrolled primary care study found that 17% of patients misused opioid medications after a treatment agreement (41). Two studies in pain specialty settings reported misuse of opioids during the year after an agreement in 24% to 28% of patients (43, 44). A study of patients in a pain specialty setting without a history of substance abuse found that only 3% misused opioids (42). The only study of opioid misuse after implementing routine urine drug testing without an opioid treatment agreement, conducted in 2 pain centers, found that 43% of patients misused opioids over nearly 4 years (45). The quality of these studies was rated as fair for 5 (39–41, 43, 45) and poor for 2 (42, 44). In addition to lack of a control group, a common quality concern is sampling bias because 2 studies excluded patients with known substance abuse (42, 43), whereas another 2 included patients who received referrals because of substance abuse concerns (39, 40).

Discussion

Despite an increasing reliance on prescribed opioids to treat chronic noncancer pain and a concurrent increase in diversion and overdose deaths (17, 18), this systematic review found that few studies have been conducted on the use of opioid treatment agreements and urine drug testing to reduce opioid misuse. Opioid misuse outcome measures varied substantially across studies, with none examining the clinically important outcomes of opioid abuse, dependence, overdose, or death. Moreover, all 11 studies used observational designs that are vulnerable to several threats to validity, such as unmeasured confounding. In the 4 studies that included comparison groups, multicomponent management strategies were associated with a 7% to 23% reduction in patient misuse of opioids when compared with preintervention conditions or historical control participants.

These studies were not representative of practice settings in which most patients are prescribed long-term opioids for chronic noncancer pain. Although we searched for studies conducted in any outpatient setting, most were conducted in pain specialty settings or in primary care practices that offered complex management strategies, including monitoring by ancillary or interdisciplinary staff. However, most patients with chronic pain are managed in

Table 2. **Studies Evaluating Opioid Misuse After Implementation of OTAs or UDT**

Study, Year (Reference)	Study Design	Setting	Sample Characteristics*	Management Strategy
Wiedemer et al, 2007 (36)	RCS	VA-based primary care with consultative support	335 patients referred by a PCP	OTA and routine UDT, with regular monitoring visits with clinical pharmacist
Goldberg et al, 2005 (35)	RCS	VA-based primary care	91 patients with an OTA in their medical record; mean age, 52 y; 88% men; 66% white	OTA‡
Manchikanti et al, 2006 (37)	PCS	Pain clinic	500 consecutive patients receiving stable opioid regimen; mean age, 48.5 y; 41% men; mean pain duration, 10.7 y	OTA and routine UDT
Manchikanti el al, 2006 (38)	PCS	Pain clinic	500 consecutive patients receiving stable opioid regimen; mean age, 48.5 y; 41% men; mean pain duration, 10.7 y	OTA and routine UDT
Chelminski et al, 2005 (39)	PCS	Primary care with consultative support	85 patients referred by a PCP to a pain program; mean age, 51 y; 60% men; 78% white; 75% had a history of alcohol use; 44% had a history of substance use	OTA, routine UDT, monthly visits in the context of a multidisciplinary pain management program††
Ives et al, 2006 (40)	PCS	Primary care with consultative support	196 patients referred by a PCP to a pain program; mean age, 52 y; 55% men; 75% white; 28% had a history of alcohol abuse; 29% had a history of cocaine abuse	OTA, routine UDT, monthly visits in the context of a multidisciplinary pain management program††
Hariharan et al, 2007 (41)	RCS	Primary care	330 patients; median age, 49 y; 52% men; 50% white; 48% black	OTA and UDT
Compton et al, 2008 (43)	PCS	VA-based pain clinic	135 patients without current substance use disorder defined by DSM-IV; mean age, 53 y; 94% men	OTA and monthly visits
Burchman and Pagel, 1995 (42)	RCS	VA-based pain clinic	81 patients with no known history of substance use, followed in clinic ≥1 y	OTA and monthly visits
Vaglienti et al, 2003 (44)	RCS	Pain clinic	Approximately 780 patients with OTA; of 184 patients with infractions, 51% were men	OTA with monthly UDT
Katz et al, 2003 (45)	RCS	Pain clinic	All 122 patients receiving long-term opioid therapy at 2 academic pain centers; mean age, 45 y; 51% men; 89% white; 17% had a history of substance abuse	Routine UDT¶¶

ARR = absolute risk reduction; BEH = behavioral monitoring (includes overuse, lost or stolen prescriptions, and provider impression of misuse); DSM-IV = *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition*; GRADE = Grading of Recommendations Assessment, Development, and Evaluation; iUDT = urine drug test results inappropriately positive for an illicit or nonprescribed drug or negative for a prescribed opioid; MS = receipt of opioid medications from multiple sources; OTA = opioid treatment agreement; PCP = primary care physician; PCS = prospective cohort study; RCS = retrospective cohort study; RxA = prescription adulteration or forgery; UDT = urine drug testing; VA = Veterans Administration; VIOL = other or unspecified violation of OTA.
* Sample characteristics provided, as available.
† Quality rating: excellent (score ≥11 and GRADE ≥ 3), good (score ≥11 and GRADE ≥2), fair (score = 6–10 and GRADE ≥2), or poor (score ≤5 or GRADE = 1).
‡ Terms of OTA included UDT.
§ Matched on age, sex, race, and duration and dose of opioid treatment.
‖ Mean number of opioid prescribers per patient per quarter >2 (yes or no) during the 1.5 y after OTA compared with 2 y before OTA.
¶ Obtaining opioids either illegally or from another physician in nonemergency circumstances (yes or no).
** UDT detected the presence of cocaine, marijuana, methamphetamines, or amphetamines.
†† Multidisciplinary team included the patient's PCP, a clinical pharmacist, program assistant, pain psychiatrist, and nurse.
‡‡ A single marijuana-positive UDT result was not included in the definition of "opioid misuse."
§§ UDT detected the presence of cocaine or marijuana.
‖‖ Attempted to fill opioid prescriptions from an outside source.
¶¶ UDT frequency varied at the 2 sites (at one, at least annually or more if the physician found it to be indicated; at the other, at approximately every visit).
*** Behaviors included lost or stolen prescriptions, opioid overuse, visits without appointments, intolerance and allergies to several drugs, and frequent telephone calls.

primary care settings (1) that probably lack the staffing and support to implement these complex strategies.

In 7 studies that did not have control groups, we observed wide variation in opioid misuse after implementation of treatment agreements, urine drug testing, or both (4% to 32% in primary care and 3% to 43% in pain clinics). This variation probably reflects differences in the underlying risk for misuse in the study samples. Hariharan and colleagues (41) examined a representative sample of primary care patients and, after treatment agreements, identified opioid misuse (for example, prescription adulteration, several opioid sources, or other opioid-related over-the-counter violations) in only 4% of patients. In comparison, a literature review by Kahan and coworkers (46) reported that 7% to 31% of primary care patients who received treatment with long-term opioid medications but without treatment agreements exhibited signs of opioid misuse, such as repeatedly requesting early refills or reporting lost or stolen medications. Unfortunately, differences in the measures of opioid misuse limit the conclusions we can draw about the effect of the agreement itself.

A recent review by Chou and colleagues (47) examined a broad range of issues about opioid management and found little support for monitoring strategies, such as opioid treatment agreements and urine drug testing, but cited only 1 relevant study. In our systematic review, we identi-

Table 2—**Continued**

Control Condition	Follow-up	Misuse Measure	Results	Limitations	Quality†
Pre–post analysis	2-y observation	iUDT and BEH	ARR, 22.9% (95% CI, 17.3%–28.7%)	Sampling bias	Fair
Pre–post analysis; also described temporal change in 224 matched patients without an OTA§	4-y observation	MS‖	ARR, 20.9% (CI, 10.6%–31.2%); no temporal change among patients without an OTA	Potential confounding intervention; misuse measure may not be clinically relevant	Fair
Historical control group from the same setting	Not provided	MS¶	ARR, 8.6% (CI, 4.4%–12.8%)	Historical control group	Poor
Historical control group from the same setting	Not provided	iUDT**	ARR, 6.5% (CI, 1.3%–11.7%)	Historical control group	Poor
None	3-mo follow-up	iUDT‡‡, MS, RxA, VIOL	27 (32%) with misuse	Sampling bias	Fair
None	1-y follow-up	iUDT‡‡, MS, RxA, VIOL	62 (32%) with misuse	Sampling bias	Fair
None	Median 22.5-mo follow-up during 5-y observation	MS, RxA, iUDT, VIOL	54 (17%) with any misuse; 14 (4%) with prescription drug abuse; 53 (16%) with illicit drug use§§	Incomplete UDT data	Fair
None	1-y observation	VIOL, which included iUDT and BEH	38 (28%) with any OTA violation; 15 (11%) with problematic opioid use	Sampling bias	Fair
None	3-y observation	MS‖ ‖	2 (3%) with misuse	Misuse measure may not be clinically relevant	Poor
None	1-y observation	iUDT	184 (approximately 24%) with misuse	Sample size was imprecise	Poor
None	3.75-y observation	iUDT and BEH***	53 (43%) with misuse	UDT frequency varied between 2 sites	Fair

fied 11 relevant studies and provide a comprehensive synthesis of the available evidence base. Our finding of limited evidence for the effectiveness of opioid treatment agreements should be viewed in the context of findings from a Cochrane review of treatment agreements (contracts) between patients and health care providers for clinical indications other than chronic pain, such as addiction, hypertension, and obesity (48). In that review, 16 of 30 randomized, controlled trials found statistically significant improvements in outcomes, including abstinence from substance abuse and adherence to medications and behavioral regimens.

Despite the lack of rigorous evidence supporting the use of treatment agreements and urine drug testing, these strategies have been endorsed by pain and addiction experts, professional societies, and regulatory agencies (19–21), and their use has been proposed as a quality indicator (49). Even in the absence of strong evidence, several compelling reasons for physicians to consider implementing these strategies exist. First, primary care providers who use opioid treatment agreements report improved satisfaction, comfort, and sense of mastery in managing chronic pain (50, 51). Second, management strategies that include treatment agreements have been associated with reductions in emergency department visits in observational studies (35, 36). Third, cross-sectional studies and a case series have demonstrated that urine drug testing is a valuable tool to detect use of nonprescribed drugs and confirm adherence to prescribed medications beyond that identified by patient self-report or impression of the treating physician (45, 52–57). Finally, implementing routine urine drug testing may improve the provider–patient relationship and clinic morale, as suggested in a letter to the editor (58).

In addition to lack of a strong evidence base, barriers to implementing treatment agreements and urine drug testing exist, particularly in primary care practices. Clinicians may be concerned about the time required to complete a treatment agreement with the patient (36) or that committing to a treatment agreement will restrict their clinical decision making. Furthermore, some clinicians may regard it as unethical to require treatment agreements for patients who take opioid analgesics but not for patients who take other potentially dangerous therapies, such as warfarin or insulin (59). Barriers to conducting urine drug testing in primary care practices include discomfort with discussing testing with patients, lack of access to appropriate tests (60), confusion about how to interpret or respond to test results (61, 62), and belief that one's patients are not at risk for opioid misuse and urine drug testing would be unnecessary. Misinterpretation of test re-

sults can lead to falsely accusing patients of opioid misuse and consequently harming the provider–patient relationship (60). Future research must address barriers and potential harms when evaluating the effectiveness of implementing these strategies.

Our review is limited by the considerable variation in the definition of opioid misuse in identified studies. Measures of opioid misuse ranged from subjective (provider impression of misuse) to the more objective finding of a positive urine drug test for an illicit substance. These measures also vary in severity; experts believe that adulterating a prescription is more likely to indicate addiction than obtaining opioids from multiple sources, which could represent undertreatment of pain (63). Unfortunately, no measure of opioid misuse has been demonstrated to be clearly superior in predicting clinically important outcomes, such as opioid abuse, dependence, overdose, or death. Research on defining optimal measures for opioid misuse is necessary because the use of clinical outcomes, such as opioid dependence, overdose, or death, would require large study populations and long time frames to identify events.

Another limitation is the multicomponent nature of most opioid management strategies, which does not distinguish the most effective components. Even among opioid treatment agreements, wide variation exists (64, 65). Experts have promoted goal-directed agreements that acknowledge the lack of evidence supporting long-term opioid treatment of chronic pain by defining functional goals to be achieved by the patient in order to continue the opioid analgesics (66, 67). In this review, we identified only 1 study that used a goal-directed agreement (42), and we could not draw any conclusions about its superiority over other approaches. Similarly, the implementation of urine drug testing interventions (for example, frequency, or type of assays) varied among studies and was reported inconsistently. Finally, publication bias and language bias may have limited the evidence available for our systematic review.

Our systematic review reveals that weak evidence supports the use of opioid treatment agreements and urine drug testing to reduce opioid misuse, despite the theoretical benefits of these strategies. This lack of evidence may explain in part why they have not been widely adopted in primary care. However, serious consequences of opioid misuse are on the rise (14–18) and affect all age groups (68, 69). Urgent attention is necessary to address this glaring deficiency in research, as are studies on the effectiveness of long-term opioid therapy for noncancer pain. Future research on effective risk-reduction strategies needs to be conducted in diverse primary care settings and use standardized measures of misuse, ideally those associated with clinical outcomes. In addition, this research should assess not only the benefits but also the potential harms of these strategies, such as patients forgoing pain treatment because of a perceived stigma or clinicians undertreating pain because of the perceived burden of opioid risk management. Studies of opioid management strategies need to be conducted as part of a larger initiative to improve the safety and effectiveness of long-term opioid treatment of patients with chronic noncancer pain.

From Albert Einstein College of Medicine and Montefiore Medical Center, Bronx, New York; Yale University School of Medicine, New Haven, Connecticut; Boston University School of Medicine and Boston Medical Center, Boston, Massachusetts; and University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania.

Preliminary results of this study were presented at the 32nd annual meeting of the Society of General Internal Medicine, Miami Beach, Florida, 13–16 May 2009.

Grant Support: Through the Program of Research Integrating Substance Use Issues into Mainstream Health Care, funded by the Robert Wood Johnson Foundation, the National Institute on Drug Abuse, and the Substance Abuse and Mental Health Services Administration. Dr. Starrels was supported in part by the Robert Wood Johnson Foundation Clinical Scholars Program.

Acknowledgment: The authors thank Karen Lillie, Carlos Rodriguez, and Moonseong Heo for contributing to the literature search and the Montefiore Division of General Internal Medicine Substance Abuse Research Group.

Potential Conflicts of Interest: Disclosures can be viewed at www.acponline.org/authors/icmje/ConflictOfInterestForms.do?msNum=M09-2037.

Requests for Single Reprints: Barbara J. Turner, MD, MSEd, Division of General Internal Medicine, University of Pennsylvania School of Medicine, 1123 Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19134; e-mail, bturner@mail.med.upenn.edu.

Current author addresses and author contributions are available at www.annals.org.

References

1. **Caudill-Slosberg MA, Schwartz LM, Woloshin S.** Office visits and analgesic prescriptions for musculoskeletal pain in US: 1980 vs. 2000. Pain. 2004;109:514-9. [PMID: 15157714]

2. **Cherry DK, Woodwell DA, Rechtsteiner EA.** National Ambulatory Medical Care Survey: 2005 summary. Adv Data. 2007:1-39. [PMID: 17703793]

3. **Clark JD.** Chronic pain prevalence and analgesic prescribing in a general medical population. J Pain Symptom Manage. 2002;23:131-7. [PMID: 11844633]

4. **Elliott AM, Smith BH, Penny KI, Smith WC, Chambers WA.** The epidemiology of chronic pain in the community. Lancet. 1999;354:1248-52. [PMID: 10520633]

5. **Gureje O, Von Korff M, Simon GE, Gater R.** Persistent pain and well-being: a World Health Organization Study in Primary Care. JAMA. 1998;280:147-51. [PMID: 9669787]

6. **Olsen Y, Daumit GL, Ford DE.** Opioid prescriptions by U.S. primary care physicians from 1992 to 2001. J Pain. 2006;7:225-35. [PMID: 16618466]

7. **Zerzan JT, Morden NE, Soumerai S, Ross-Degnan D, Roughead E, Zhang F, et al.** Trends and geographic variation of opiate medication use in state Medicaid fee-for-service programs, 1996 to 2002. Med Care. 2006;44:1005-10. [PMID: 17063132]

8. **IMS Health.** 2008 Top Therapeutic Classes by U.S. Dispensed Prescriptions. Accessed at www.imshealth.com/portal/site/imshealth/menuitem.a46c6d4df3db4b3d88f611019418c22a/?vgnextoid=85f4a56216a10210VgnV

CM100000ed152ca2RCRD&cpsextcurrchannel=1 on 16 April 2010.
9. **Portenoy RK, Foley KM.** Chronic use of opioid analgesics in non-malignant pain: report of 38 cases. Pain. 1986;25:171-86. [PMID: 2873550]
10. **Martell BA, O'Connor PG, Kerns RD, Becker WC, Morales KH, Kosten TR, et al.** Systematic review: opioid treatment for chronic back pain: prevalence, efficacy, and association with addiction. Ann Intern Med. 2007;146:116-27. [PMID: 17227935]
11. **Ballantyne JC, Mao J.** Opioid therapy for chronic pain. N Engl J Med. 2003;349:1943-53. [PMID: 14614170]
12. **Eisenberg E, McNicol ED, Carr DB.** Efficacy and safety of opioid agonists in the treatment of neuropathic pain of nonmalignant origin: systematic review and meta-analysis of randomized controlled trials. JAMA. 2005;293:3043-52. [PMID: 15972567]
13. **Noble M, Tregear SJ, Treadwell JR, Schoelles K.** Long-term opioid therapy for chronic noncancer pain: a systematic review and meta-analysis of efficacy and safety. J Pain Symptom Manage. 2008;35:214-28. [PMID: 18178367]
14. **Substance Abuse and Mental Health Services Administration**. Results from the 2007 National Survey on Drug Use and Health Report: Patterns and Trends in Nonmedical Prescription Pain Reliever Use: 2002 to 2005. Rockville, MD: Office of Applied Studies; 2008. NSDUH Series H-34, DHHS publication no. SMA 08-4343. Accessed at www.oas.samhsa.gov/2k7/pain/pain.pdf on 20 April 2010.
15. **Substance Abuse and Mental Health Services Administration**. Treatment Episode Data Set (TEDS) Highlights—2006 National Admissions to Substance Abuse Treatment Services. Rockville, MD: U.S. Department of Health and Human Services; 2007. OAS Series #S-40, DHHS publication no. SMA 08-4313.
16. **Drug Abuse Warning Network.** The DAWN Report. National Estimates of Drug-Related Emergency Department Visits. Substance Abuse and Mental Health Services Administration. Rockville, MD: Substance Abuse and Mental Health Services Administration. Accessed at http://dawninfo.samhsa.gov/files/ED2006/DAWN2k6ED.pdf on 20 April 2010.
17. **Centers for Disease Control and Prevention (CDC).** Unintentional poisoning deaths—United States, 1999-2004. MMWR Morb Mortal Wkly Rep. 2007;56:93-6. [PMID: 17287712]
18. **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. [PMID: 15276196]
19. The use of opioids for the treatment of chronic pain. A consensus statement from the American Academy of Pain Medicine and the American Pain Society. Clin J Pain. 1997;13:6-8. [PMID: 9084947]
20. **Chou R, Fanciullo GJ, Fine PG, Adler JA, Ballantyne JC, Davies P, et al; American Pain Society-American Academy of Pain Medicine Opioids Guidelines Panel.** Clinical guidelines for the use of chronic opioid therapy in chronic noncancer pain. J Pain. 2009;10:113-30. [PMID: 19187889]
21. Model Policy for the Use of Controlled Substances for the Treatment of Pain. Federation of State Medical Boards of the United States; 2004. Accessed at www.fsmb.org/pdf/2004_grpol_Controlled_Substances.pdf on 25 April 2010.
22. **Gourlay DL, Heit HA, Almahrezi A.** Universal precautions in pain medicine: a rational approach to the treatment of chronic pain. Pain Med. 2005;6:107-12. [PMID: 15773874]
23. **Adams NJ, Plane MB, Fleming MF, Mundt MP, Saunders LA, Stauffacher EA.** Opioids and the treatment of chronic pain in a primary care sample. J Pain Symptom Manage. 2001;22:791-6. [PMID: 11532592]
24. **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;22:1859-65. [PMID: 16968589]
25. **Boulanger A, Clark AJ, Squire P, Cui E, Horbay GL.** Chronic pain in Canada: have we improved our management of chronic noncancer pain? Pain Res Manag. 2007;12:39-47. [PMID: 17372633]
26. **Linzer M, Manwell LB, Williams ES, Bobula JA, Brown RL, Varkey AB, et al; MEMO (Minimizing Error, Maximizing Outcome) Investigators.** Working conditions in primary care: physician reactions and care quality. Ann Intern Med. 2009;151:28-36, W6-9. [PMID: 19581644]
27. **Bair MJ.** Overcoming fears, frustrations, and competing demands: an effective integration of pain medicine and primary care to treat complex pain patients [Editorial]. Pain Med. 2007;8:544-5. [PMID: 17883738]
28. **Willi C, Bodenmann P, Ghali WA, Faris PD, Cornuz J.** Active smoking and the risk of type 2 diabetes: a systematic review and meta-analysis. JAMA. 2007;298:2654-64. [PMID: 18073361]
29. **Ballantyne JC, LaForge KS.** Opioid dependence and addiction during opioid treatment of chronic pain. Pain. 2007;129:235-55. [PMID: 17482363]
30. **Downs SH, Black N.** The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. J Epidemiol Community Health. 1998;52:377-84. [PMID: 9764259]
31. **Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, et al.** Assessing the quality of reports of randomized clinical trials: is blinding necessary? Control Clin Trials. 1996;17:1-12. [PMID: 8721797]
32. **Fung CH, Lim YW, Mattke S, Damberg C, Shekelle PG.** Systematic review: the evidence that publishing patient care performance data improves quality of care. Ann Intern Med. 2008;148:111-23. [PMID: 18195336]
33. **Atkins D, Briss PA, Eccles M, Flottorp S, Guyatt GH, Harbour RT, et al; GRADE Working Group.** Systems for grading the quality of evidence and the strength of recommendations II: pilot study of a new system. BMC Health Serv Res. 2005;5:25. [PMID: 15788089]
34. **Higgins JP, Green S, eds.** Cochrane Handbook for Systematic Reviews of Interventions, Version 5.0.1 (updated September 2008). The Cochrane Collaboration; 2008. Accessed at www.cochrane-handbook.org on 20 April 2010.
35. **Goldberg KC, Simel DL, Oddone EZ.** Effect of an opioid management system on opioid prescribing and unscheduled visits in a large primary care clinic. J Clin Outcomes Management. 2005;12:621-28.
36. **Wiedemer NL, Harden PS, Arndt IO, Gallagher RM.** The opioid renewal clinic: a primary care, managed approach to opioid therapy in chronic pain patients at risk for substance abuse. Pain Med. 2007;8:573-84. [PMID: 17883742]
37. **Manchikanti L, Manchukonda R, Damron KS, Brandon D, McManus CD, Cash K.** Does adherence monitoring reduce controlled substance abuse in chronic pain patients? Pain Physician. 2006;9:57-60. [PMID: 16700282]
38. **Manchikanti L, Manchukonda R, Pampati V, Damron KS, Brandon DE, Cash KA, et al.** Does random urine drug testing reduce illicit drug use in chronic pain patients receiving opioids? Pain Physician. 2006;9:123-9. [PMID: 16703972]
39. **Chelminski PR, Ives TJ, Felix KM, Prakken SD, Miller TM, Perhac JS, et al.** A primary care, multi-disciplinary disease management program for opioid-treated patients with chronic non-cancer pain and a high burden of psychiatric comorbidity. BMC Health Serv Res. 2005;5:3. [PMID: 15649331]
40. **Ives TJ, Chelminski PR, Hammett-Stabler CA, Malone RM, Perhac JS, Potisek NM, et al.** Predictors of opioid misuse in patients with chronic pain: a prospective cohort study. BMC Health Serv Res. 2006;6:46. [PMID: 16595013]
41. **Hariharan J, Lamb GC, Neuner JM.** Long-term opioid contract use for chronic pain management in primary care practice. A five year experience. J Gen Intern Med. 2007;22:485-90. [PMID: 17372797]
42. **Burchman SL, Pagel PS.** Implementation of a formal treatment agreement for outpatient management of chronic nonmalignant pain with opioid analgesics. J Pain Symptom Manage. 1995;10:556-63. [PMID: 8537698]
43. **Compton PA, Wu SM, Schieffer B, Pham Q, Naliboff BD.** Introduction of a self-report version of the Prescription Drug Use Questionnaire and relationship to medication agreement noncompliance. J Pain Symptom Manage. 2008;36:383-95. [PMID: 18508231]
44. **Vaglienti RM, Huber SJ, Noel KR, Johnstone RE.** Misuse of prescribed controlled substances defined by urinalysis. W V Med J. 2003;99:67-70. [PMID: 12874916]
45. **Katz NP, Sherburne S, Beach M, Rose RJ, Vielguth J, Bradley J, et al.** Behavioral monitoring and urine toxicology testing in patients receiving long-term opioid therapy. Anesth Analg. 2003;97:1097-102. [PMID: 14500164]
46. **Kahan M, Srivastava A, Wilson L, Gourlay D, Midmer D.** Misuse of and dependence on opioids: study of chronic pain patients. Can Fam Physician. 2006;52:1081-7. [PMID: 17279218]
47. **Chou R, Fanciullo GJ, Fine PG, Miaskowski C, Passik SD, Portenoy RK.** Opioids for chronic noncancer pain: prediction and identification of aberrant drug-related behaviors: a review of the evidence for an American Pain Society and American Academy of Pain Medicine clinical practice guideline. J Pain. 2009;10:131-46. [PMID: 19187890]
48. **Bosch-Capblanch X, Abba K, Prictor M, Garner P.** Contracts between patients and healthcare practitioners for improving patients' adherence to treatment, prevention and health promotion activities. Cochrane Database Syst Rev. 2007:CD004808. [PMID: 17443556]
49. **Institute for Clinical Systems Improvement.** Assessment and management of chronic pain: percentage of patients diagnosed with chronic pain who are pre-

scribed an opioid who have an opioid agreement form and urine toxicology screen documented in the medical record (July 2008). Accessed at www.qualitymeasures.ahrq.gov/summary/summary.aspx?doc_id=13005&string=urine+AND+pain on 16 April 2010.
50. **Fagan MJ, Chen JT, Diaz JA, Reinert SE, Stein MD.** Do internal medicine residents find pain medication agreements useful? Clin J Pain. 2008;24:35-8. [PMID: 18180634]
51. **Touchet BK, Yates WR, Coon KA.** Opioid contract use is associated with physician training level and practice specialty. J Opioid Manag. 2005;1:195-200. [PMID: 17315546]
52. **Fleming MF, Balousek SL, Klessig CL, Mundt MP, Brown DD.** Substance use disorders in a primary care sample receiving daily opioid therapy. J Pain. 2007;8:573-82. [PMID: 17499555]
53. **Fishbain DA, Cutler RB, Rosomoff HL, Rosomoff RS.** Validity of self-reported drug use in chronic pain patients. Clin J Pain. 1999;15:184-91. [PMID: 10524471]
54. **Atluri S, Sudarshan G.** Evaluation of abnormal urine drug screens among patients with chronic non-malignant pain treated with opioids. Pain Physician. 2003;6:407-9. [PMID: 16871289]
55. **Berndt S, Maier C, Schütz HW.** Polymedication and medication compliance in patients with chronic non-malignant pain. Pain. 1993;52:331-9. [PMID: 8460051]
56. **Manchikanti L, Damron KS, Beyer CD, Pampati V.** A comparative evaluation of illicit drug use in patients with or without controlled substance abuse in interventional pain management. Pain Physician. 2003;6:281-5. [PMID: 16880872]
57. **Schuckman H, Hazelett S, Powell C, Steer S.** A validation of self-reported substance use with biochemical testing among patients presenting to the emergency department seeking treatment for backache, headache, and toothache. Subst Use Misuse. 2008;43:589-95. [PMID: 18393078]
58. **Sampson JM, Acholonlu WW Jr.** Reducing patient aggression and hostility in primary care with urine drug testing [Letter]. South Med J. 2004;97:916-7. [PMID: 15455991]
59. **Gitlin MC.** Contracts for opioid administration for the management of chronic pain: a reappraisal [Editorial]. J Pain Symptom Manage. 1999;18:6-8. [PMID: 10439567]
60. **Heit HA, Gourlay DL.** Urine drug testing in pain medicine. J Pain Symptom Manage. 2004;27:260-7. [PMID: 15010104]
61. **Reisfield GM, Webb FJ, Bertholf RL, Sloan PA, Wilson GR.** Family physicians' proficiency in urine drug test interpretation. J Opioid Manag. 2007;3:333-7. [PMID: 18290585]
62. **Tellioglu T.** The use of urine drug testing to monitor patients receiving chronic opioid therapy for persistent pain conditions. Med Health R I. 2008;91:279-80, 282. [PMID: 18834046]
63. **Fine PG, Portenoy RK.** A Clinical Guide to Opioid Analgesia. Minneapolis: McGraw-Hill; 2004.
64. **Fishman SM, Bandman TB, Edwards A, Borsook D.** The opioid contract in the management of chronic pain. J Pain Symptom Manage. 1999;18:27-37. [PMID: 10439570]
65. **Roskos SE, Keenum AJ, Newman LM, Wallace LS.** Literacy demands and formatting characteristics of opioid contracts in chronic nonmalignant pain management. J Pain. 2007;8:753-8. [PMID: 17382596]
66. **McCleane G, Smith HS.** Opioids for persistent noncancer pain. Anesthesiol Clin. 2007;25:787-807, vi-ii. [PMID: 18054145]
67. **Ballantyne JC.** Opioids for chronic nonterminal pain. South Med J. 2006;99:1245-55. [PMID: 17195420]
68. **Sung HE, Richter L, Vaughan R, Johnson PB, Thom B.** Nonmedical use of prescription opioids among teenagers in the United States: trends and correlates. J Adolesc Health. 2005;37:44-51. [PMID: 15963906]
69. **Pergolizzi J, Böger RH, Budd K, Dahan A, Erdine S, Hans G, et al.** Opioids and the management of chronic severe pain in the elderly: consensus statement of an International Expert Panel with focus on the six clinically most often used World Health Organization Step III opioids (buprenorphine, fentanyl, hydromorphone, methadone, morphine, oxycodone). Pain Pract. 2008;8:287-313. [PMID: 18503626]

Current Author Addresses: Dr. Starrels: Division of General Internal Medicine, Albert Einstein College of Medicine and Montefiore Medical Center, 111 East 210th Street, Bronx, NY 10467.
Dr. Becker: Section of General Internal Medicine, Yale University School of Medicine, 367 Cedar Street, PO Box 208093, New Haven, CT 06520-8093.
Dr. Alford: Section of General Internal Medicine, Clinical Addiction, Research and Education Unit, Boston University School of Medicine and Boston Medical Center, 801 Massachusetts Avenue, 2nd Floor, Boston, MA 02118.
Dr. Kapoor: Section of General Internal Medicine, Boston University School of Medicine and Boston Medical Center, 801 Massachusetts Avenue, 2nd Floor, Boston, MA 02118.
Mr. Williams: 4229 Regent Square, Philadelphia, PA 19104.
Dr. Turner: Division of General Internal Medicine, University of Pennsylvania School of Medicine, 1123 Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19134.

Author Contributions: Conception and design: J.L. Starrels, W.C. Becker, D.P. Alford, A. Kapoor.
Analysis and interpretation of the data: J.L. Starrels, W.C. Becker, D.P. Alford, A. Kapoor, A.R. Williams, B.J. Turner.
Drafting of the article: J.L. Starrels, D.P. Alford, A. Kapoor, A.R. Williams, B.J. Turner.
Critical revision of the article for important intellectual content: J.L. Starrels, W.C. Becker, D.P. Alford, A. Kapoor, B.J. Turner.
Final approval of the article: J.L. Starrels, W.C. Becker, D.P. Alford, A.R. Williams, B.J. Turner.
Statistical expertise: J.L. Starrels, W.C. Becker.
Obtaining of funding: B.J. Turner.
Administrative, technical, or logistic support: A.R. Williams, B.J. Turner.
Collection and assembly of data: J.L. Starrels, W.C. Becker, D.P. Alford, A.R. Williams, B.J. Turner.

Appendix Table 1. **Search Terms***

Themes	MeSH Terms†	Keywords or Text Words‡
1. Opioid medications	Narcotics, analgesics, opioid	opi$, narcotic$, buprenorphine, butorphanol, codeine, dihydromorphine, fentanyl, heroin, hydrocodone, hydromorphone, levorphanol, meperidine, methadone, morphine, nalbuphine, oxycodone, oxymorphone, propoxyphene
2. Chronic pain	Pain	pain$
	Chronic disease	chronic pain, pain management, noncancer pain, pain syndrome, pain treatment, pain control, nonmalignant pain
3a. Opioid treatment agreements	Contracts	agreement$, contract$, treatment plan, informed consent
3b. Urine drug testing	Substance abuse detection	drug test$, drug screen$, drug monitor$, urine test$, urine screen$, toxicology
	Mass screening	screen$, opi$ assess$, addict$ screen$

MeSH = Medical Subject Heading.
* MeSH terms and keywords and text words within each theme were combined by using the Boolean operator "or." Themes were then combined by using the Boolean operator "and" as (theme 1) "and" (theme 2) "and" (theme 3a or theme 3b).
† MEDLINE, EMBASE, and Cochrane Central Register of Controlled Trials.
‡ Keywords were used in MEDLINE and EMBASE, and text words were used in Cochrane Central Register of Controlled Trials and PsycINFO.

Appendix Table 2. **Quality of Studies Assessing Opioid Misuse Outcomes After Opioid Treatment Agreements or Urine Drug Testing**

Domain	Study, Year (Reference)										
	Wiedemer et al, 2007 (36)	Goldberg et al, 2005 (35)	Manchikanti et al, 2006 (37)	Manchikanti et al, 2006 (38)	Chelminski et al, 2005 (39)	Ives et al, 2006 (40)	Hariharan et al, 2007 (41)	Compton et al, 2008 (43)	Burchman and Pagel, 1995 (42)	Vaglienti et al, 2003 (44)	Katz et al, 2003 (45)
Sample domain											
Persons with diverse pain types included	√	√	√	√	√	√	√	√	√	√	√
Persons with history of substance abuse included*	√	√	√	√	√	√	√	0	0	√	√
Persons with mental health disorders included*	√	√	√	√	√	√	√	√	√	√	√
Score (maximum 3 points)	3	3	3	3	3	3	3	2	2	3	3
Design domain†											
Prospective design	0	0	√	√	√	√	0	√	0	0	0
Control group included	√	√	√	√	0	0	0	0	0	0	0
Control participants from a similar population	√	√	0	0	0	0	0	0	0	0	0
Intervention described clearly	√	√	0	√	√	√	√	0	√	0	0
Intervention consistent among groups	0	0	0	0	√	0	0	0	√	0	0
Outcome described clearly	√	√	0	√	√	√	√	√	0	√	√
Outcome objective	√	√	0	√	√	√	√	√	0	√	√
Completion or response rate ≥85%	0	√	√	√	0	√	√	√	√	√	√
Distribution of potential confounders provided	0	0	0	0	0	0	0	0	0	0	0
Multivariate analysis conducted	0	0	0	0	0	√	√	0	0	0	0
Adequate adjustment for confounding	0	0	0	0	0	0	0	0	0	0	0
Results clearly presented	√	√	0	√	√	√	√	√	0	0	√
Score (maximum 12 points)	6	7	3	7	6	7	6	5	3	3	4
Sample + design score (maximum 15 points)	9	10	6	10	9	10	9	7	5	6	7
GRADE score (maximum 4 points)‡	2	2	1	1	2	2	2	2	2	1	2
Quality rating	Fair	Fair	Poor	Poor	Fair	Fair	Fair	Fair	Poor	Poor	Fair

GRADE = Grading of Recommendations Assessment, Development, and Evaluation.
* Studies were scored with a check if patients with substance abuse or mental health disorders were not explicitly excluded.
† Quality items were scored as 0 when they were not applicable to the study design.
‡ See **Appendix Table 3** for derivation of GRADE score.

Appendix Table 3. **Derivation of GRADE Score**

Study, Year (Reference)	Initial GRADE Score*	Adjustment†	Reason for Adjustment	Final GRADE Score
Wiedemer et al, 2007 (36)	2	–	–	2
Goldberg et al, 2005 (35)	2	–	–	2
Manchikanti et al, 2006 (37)	2	−1	Selection bias‡	1
Manchikanti et al, 2006 (38)	2	−1	Selection bias‡	1
Chelminski et al, 2005 (39)	2	–	–	2
Ives et al, 2006 (40)	2	–	–	2
Hariharan et al, 2007 (41)	2	–	–	2
Compton et al, 2008 (43)	2	–	–	2
Burchman and Pagel, 1995 (42)	2	–	–	2
Vaglienti et al, 2003 (44)	2	−1	Missing data§	1
Katz et al, 2003 (45)	2	–	–	2

GRADE = Grading of Recommendations Assessment, Development, and Evaluation.
* The initial GRADE score was based on study design: 4 = randomized trial, 3 = quasi-randomized trial, 2 = observational study, 1 = any other evidence (33).
† GRADE score was decreased for quality limitations or increased for strong association (33).
‡ Comparison with a historical control group that was not well described and differed from the intervention group in inclusion criteria and in exposure to contaminating co-interventions.
§ The precise number of participants receiving the intervention was not reported.

Annals of Internal Medicine | PERSPECTIVE

Concierge Medicine: A "Regular" Physician's Perspective

Michael Stillman, MD

Concierge medical practices, which advertise expanded access to care and individualized attention, collect charges both from insurance companies and directly from their patients. Some bill hundreds of dollars for one-time "executive" physicals, whereas others have patients pay annual retainer fees. Yet, virtually no data are available about these "luxury" practices. It is not known how many physicians have "turned concierge," whether they have altered their testing and prescribing patterns, or whether their clinical outcomes are superior to those of their colleagues in traditional practices. Although some have voiced concern that concierge physicians create a 2-tiered system and may contribute to the difficulty that many patients have with access to care, the medical community has largely remained silent on the matter. The mere existence of concierge medicine may reflect our need as physicians to do better by our patients. Yet our responsibility as a professional community is to engage in—not run from—that monumental challenge.

Ann Intern Med. 2010;152:391-392. www.annals.org
For author affiliation, see end of text.

Late last week, Marc, a fit, middle-aged patient, sent an e-mail telling me he was joining a concierge practice. Marc, his wife, and several of his siblings—all patients in my office for nearly 20 years—had always received same-day attention for any concerns, promptly returned calls, evidence-based and ethical care, and warm service from my staff. My relationship with him had at times seemed as much friendly as professional, and he had once invited me to his vacation home for dinner. So what precipitated this breakup?

Several weeks previously, 1 day before departing for vacation, Marc had called to tell me he felt something "stuck in his throat." He denied fevers, chills, or night sweats; had no new lumps in his neck; and had no difficulty swallowing food or breathing. Apart from this minor complaint, in fact, he felt perfectly well. Although Marc did not want to come to the office, he wondered what I thought might be wrong. I told him that he probably had postnasal drip or acid reflux, but that we could certainly arrange for neck ultrasonography and, if the result was negative, laryngoscopy. I called his local hospital, scheduled the ultrasonography for the day he returned from vacation, and called Marc to let him know.

Two weeks later, I called Marc and left a message inquiring whether he had had the ultrasonography, because I had not yet seen a report. In response, Marc penned the following e-mail: "I want you to know," he wrote, "that I have enormous regard for your capabilities as a doctor . . . and have always felt you were extremely caring and thorough. It is simply," he continued, "that I have a slight tendency toward obsessive worrying, and it has been my experience that waiting in line in the world of medicine can be torture." Marc told me that when he found he could not obtain same-day ultrasonography, one of his friends had directed him to a concierge practice. He had paid an $1800 fee, been seen within an hour, gone directly for an endoscopy, and received a diagnosis of . . . acid reflux. Marc had received the prompt diagnosis he felt he needed and had therefore decided to stay with his new physician.

This is not the first time a patient has left me for a concierge practice.

Gary, a healthy man 44 years of age, called to tell me he had subacute hand and arm numbness, and I offered him a same-day visit. His symptoms and physical examination suggested ulnar neuropathy—a peripheral nerve entrapment—but one of his nonmedical friends had convinced him he had a bulging disk and needed an imaging study. I scheduled Gary for a confirmatory nerve study, and only realized he had switched physicians when he did not attend. Disappointed with my reluctance to order magnetic resonance imaging, Gary found a more accommodating "luxury" physician who had done so promptly.

Mary, a pleasant woman 60 years of age, told me at our first visit that she had lupus. She had been hospitalized several times for "autoimmune flares," received immunomodulators intermittently, and had become so accustomed to her diagnosis that she "blamed all her sickness on it." After closely reviewing Mary's records, I found no evidence of an autoimmune disorder, but when I told her this good news and offered her a reevaluation from my favorite rheumatologist, she left me for a concierge physician. Rather than being overjoyed, she felt I had "pulled a rug out from under her," and in transferring physicians cited her need for "more personal" care.

Tim, a delightful man 21 years of age, presented with lymphadenopathy, weight loss, and night sweats. At our first visit, I admitted him to my service, arranged for an expedited metastatic workup, and unfortunately diagnosed lymphoma. Although I visited Tim several times after work hours and communicated closely with his oncology team, his parents enrolled him with a concierge physician who charges a $5000 annual "membership fee." Tim's father wrote that although he appreciated my efforts and attention, he was looking for "more proactive care." I have wondered since whether Tim's father had chosen that uniquely American approach of throwing money at a problem to buy a solution.

Despite the recent explosion of luxury practices, very little is known about them. No reliable statistics report the number of physicians who have "gone concierge" (or how many patients these transitions have displaced), and their clinical outcomes remain uninvestigated. A single study in

2005 (1) reported that patients at concierge practices are less ethnically diverse than the general population and have lower rates of diabetes, heart disease, and hypertension.

Several authors have expressed their ethical misgivings over concierge medicine in well-written articles. Donohoe (2) questions how physicians trained with public monies could limit their practices to a wealthy few and argues that associations between academic medical centers and "executive" practices undermine our efforts to produce clinically and morally thoughtful trainees. Carnahan (3) suggests that concierge physicians are cherry-picking the wealthiest and healthiest patients, "leaving behind a comparatively sick population to be absorbed by others in the community." Brennan (4), however, argues that the "luxury style of practice" allows physicians to spend ample time with patients and to "undertake the ethical responsibility to put [their] welfare above everything else." Kirkpatrick (5) describes how his retainer practice—which grosses $2.5 million per year—has satisfied patients, physicians, and medical center leadership and staunched an outflow of "*important patron patients*" (italics mine) who favor the "high-service benefits of belonging to a concierge practice."

It is understandable that patients want the best care possible, and practices that advertise 24-hour physician availability, lengthy appointments, and expedited access to subspecialists and procedures are naturally appealing. Marc, perturbed by having to wait a reasonable amount of time for a nonemergent evaluation, felt gratified to receive immediate care, whereas Gary, convinced that he needed magnetic resonance imaging, must have felt his new physician was doing "everything possible" to evaluate his symptoms.

I am less charitable, however, toward concierge physicians and am surprised by the neutrality with which the medical community has addressed their work. First, each of us has vowed to "treat without exception all who seek [our] ministrations," and limiting one's practice to several hundred wealthy patients undermines this tenet of our profession. Even though economic realities and schedule limitations dictate that some physicians maintain a certain payer mix or eventually close their panels to new patients, I am certain that the legendary physicians of our profession would be embarrassed by the criteria some of our colleagues have used in selecting which patients they will and will not see.

Second, Donohoe (2) correctly asserts that many retainer practices use a "buffet approach to diagnosis which makes a mockery of evidence-based medical care." A patient with a "lump in his throat" without dysphagia need not be sent for same-day ultrasonography and a man with hand numbness, no neck pain, and a negative Spurling test result does not require cervical magnetic resonance imaging.

Third, without proof that concierge medicine is clinically superior to traditional care, any suggestion that ill patients will be more tightly managed or that they will live longer or healthier lives is specious. Some patients may be willing to pay retainer fees simply for expanded access to their physicians or for receiving their care in a well-appointed office. Yet until data demonstrate that the longer visits, "executive" physicals, or annual ancillary testing offered by so many luxury practices yield better clinical outcomes, no one should not be allowed or led to believe that prompt or expensive care is necessarily the best.

Patients deserve and desire our fullest attention and consideration. They want to be listened to, to feel that we understand and appreciate their concerns, and to know that we are their staunchest advocates. Perhaps the emergence of concierge medicine will remind us of these tenets and force us to redouble our professional efforts. Broad adoption of the "medical home" model, which emphasizes the physician–patient relationship and enhanced access to care, may one day obviate luxury care. This being said, physicians who opt out of the current system by expending their energies catering to "patron patients" rather than helping reform a deeply flawed health care system or energizing a beleaguered professional community should reenter the fray. There are patients to be cared for, both wealthy and underprivileged; ideas to be proposed and exchanged; and policymakers to be educated and persuaded. These difficult times call for engagement, not isolation and retreat.

From Boston University School of Medicine, Brookline, Massachusetts.

Potential Conflicts of Interest: None disclosed. Forms can be viewed at www.acponline.org/authors/icmje/ConflictOfInterestForms.do?msNum=M09-2107.

Requests for Single Reprints: Michael Stillman, MD, 209 Harvard Street, Suite 407, Brookline, MA 02446; e-mail, michael.stillman@bmc.org.

References

1. **Alexander GC, Kurlander J, Wynia MK.** Physicians in retainer ("concierge") practice. A national survey of physician, patient, and practice characteristics. J Gen Intern Med. 2005;20:1079-83. [PMID: 16423094]
2. **Donohoe M.** Luxury primary care, academic medical centers, and the erosion of science and professional ethics. J Gen Intern Med. 2004;19:90-4. [PMID: 14748866]
3. **Carnahan SJ.** Concierge medicine: legal and ethical issues. J Law Med Ethics. 2007;35:211-5. [PMID: 17341229]
4. **Brennan TA.** Luxury primary care—market innovation or threat to access? N Engl J Med. 2002;346:1165-8. [PMID: 11948279]
5. **Kirkpatrick J.** Concierge medicine gaining ground. Competition forces medical center into 'boutique' business. Physician Exec. 2002;28:24-6. [PMID: 12416377]

For Those on Whom We Rely

FIVE MONTHS AGO DEB—MY OFFICE MANAGER—developed pain in her right armpit. It was initially mild, yet over a course of weeks it had become severe enough that she had difficulties with typing and filing. I never pried nor insisted that she be seen by her physician, and Deb is so accustomed to minimizing her own concerns—so reluctant to "bother a doctor"—that she did not seek formal consultation until quite recently. One month ago she visited her primary care physician and was prescribed a ten-day course of ibuprofen. Two weeks ago she saw a physiatrist, who ordered magnetic resonance imaging—a study that was ultimately canceled.

Deb and I have a tacit understanding that I am not to meddle in her medical management; she deserves her privacy, and she and I are so close that my interventions would be subjective and therefore faulty. There is simply no way I could offer her my most clear-headed opinion. When her pain peaked, however, and she wondered aloud how she would get through the day, I quickly palpated her armpit and shoulder, found that no area was unusually tender, and sent her around the corner for a chest film. By the time Deb returned to our office, a radiologist had called me with unwelcome news. The film revealed a large right upper lobe mass, most likely with pleural extension.

Overriding Deb's objections, I canceled our patients for the day, brought her to our hospital's emergency department where we were met by her family, and had her admitted to the medicine service. During a two-day hospital course, she underwent a transthoracic biopsy, had both brain magnetic resonance imaging and a PET scan, and was diagnosed with squamous cell carcinoma. She returned to work the day after discharge, feeling guilty about a stack of paperwork she had left undone, but during her brief absence a pall hung over the office. It was not simply that she was not there; she does occasionally take vacation. Rather, it was that my staff and I had a jolting glimpse into what it would be like were Deb to become too ill to continue working with us. We felt amputated from some essential source of strength and goodness. We felt diminished and dull.

My first job after completing residency was at a community hospital. Eager for general internists, the chief medical officer had hired me on a guaranteed salary, given me an advertising budget with which to attract new patients, and placed me in an office with a well-established local internist. Deb was the lead secretary in that office, and she adopted me as if I were her own son. During my first weeks in practice, I knew very few of the physicians and staff who would prove essential to my ability to attract patients. Deb escorted me around the hospital, introduced me to her friends and colleagues, and had lunch with me every day so I would not be lonely. She was so thorough in handling administrative tasks, so hard-working, and so unusually sensitive in her dealings with patients, I later begged her to leave a position she had held for more than 20 years in favor of managing my own practice. The day she agreed to join me was one of the most important of my life.

Deb and I have shared many humorous moments. We have bantered with our favorite patients and giggled over how awkwardly our medical students behave (then marveled at how quickly they mature). We also share interests outside work and have compared notes on play-off games, disastrous runway couture, and the previous night's performances on *Glee* and *American Idol*. Deb was once so horrified to hear that a patient had disappeared into our private office bathroom to collect a sample for semen analysis, she dubbed the room the "Cave of Iniquity." Many years later, she and I still speak of "The Cave," and our new hires never know to what we are referring.

We have also shared a good number of tragedies. Deb and I have jointly attended patients' wakes, supported people who have struggled after a spouse's or child's death, and written sympathy letters to families of patients for whom we once cared. We have seen one another though personal and family troubles with illness, addiction, and crumbling relationships and have shared, in the first hour of our workdays before our staff and patients arrive, a great many sensitive revelations.

Deb has been my stout pillar of support—an insightful woman who will unobtrusively correct me when I am wrong and who will stand with me, even when others desert, when she feels I am right. Deb proofreads and edits my journal submissions and most sensitive correspondences before I mail them, often altering and improving on both tone and content. She seems to know exactly what I want to say, yet words my thoughts more gracefully and polishes my pieces' rough edges. Once, when I was just months out of residency, I became angry at and hung up on a woman who had had me paged and awakened at 3 o'clock in the morning for a mild cough. The woman's family threatened me—both physically and legally—and several hospital administrators called me in for a disciplinary conference. Deb's pithy response: "I would never wake someone up for a cough."

A Piece of My Mind Section Editor: Roxanne K. Young, Associate Senior Editor.

Finally, Deb has been the staunchest of patient advocates, treating them, as we all should, as if they were family. Her formal job description is simply to manage our office's finances and staffing, yet she has found struggling patients transitional housing, enrolled elderly patients in publicly funded transportation programs, arranged for nursing visits for people without adequate medical support at home, and argued down insurance companies that have wrongly billed our patients or inappropriately denied payment for our services. Patients often thank me for my care and tell me I am unusual in that I see to both their medical and nonmedical needs. I deflect this praise toward Deb. She is our patients' first point of contact, their guide through an often-baffling health care system, and their unyielding defender who takes upon herself the burden of setting injustices right.

This essay is not a remembrance. Deb, her family, and I know she has difficult months and years ahead, yet we both hope and expect that she will emerge from these menacing times and retake her place as the enthusiastic and essential hub about which so many lives revolve. Rather, I wrote this in appreciation specifically of a woman who has quietly and with extraordinary humility made my career, and more broadly about the people on whom we physicians rely.

In practices large and small, general and subspecialty, and in urban, suburban, and rural settings, some fortunate physicians' careers blossom based largely on the efforts of our managers, nurses, secretaries, and medical assistants. These people who can never be adequately compensated for their energy, enthusiasm, and investment in our work are not diagnosticians yet are every bit as important to our ill patients and their loved ones as we are. They are our public faces, projecting and amplifying only our best intentions. They are our soundings boards, our moral compasses, and our confidants. They are our part-time "spouses," coming to know us so well they can read and manage our moods at least as deftly as can our own families. They become our dear and trusted friends.

Michael D. Stillman, MD
Brookline, Massachusetts
michael.stillman@bmc.org

Acknowledgment: The author thanks Deb for permission to share her story.

> The spirit of liberty is the spirit which is not too sure that it is right; the spirit of liberty is the spirit which seeks to understand the minds of other men and women; the spirit of liberty is the spirit which weighs their interests alongside his own without bias.
>
> —Learned Hand (1872-1961)

Violence and Victims, Volume 25, Number 5, 2010

Intimate Partner Aggression Perpetration in Primary Care Chronic Pain Patients

Casey Taft, PhD
VA Boston Healthcare System, and
Boston University School of Medicine

Sonia Schwartz, BA
Boston Medical Center, and
Boston University School of Medicine

Jane M. Liebschutz, MD, MPH
Boston Medical Center, Boston University School of Medicine, and
Boston University School of Public Health

This study examined the prevalence and correlates of partner aggression perpetration in 597 primary care chronic pain patients. Approximately 30% of participants reported perpetrating low-level aggression, 12% reported injuring their partner, and 5% reported engaging in sexual coercion. Women reported more low-level aggression perpetration than men, and men reported more engagement in sexual coercion than women. Substance use disorders (SUD) were associated with all outcomes, and both aggression victimization and lifetime ratings of posttraumatic stress disorder (PTSD) were associated with low-level aggression and injuries. In multivariate analyses, gender, aggression victimization, PTSD, and SUD evidenced associations with one or more outcomes. Findings indicate a need for aggression screening in this population and highlight avenues for intervention.

Keywords: chronic pain; aggression; primary health care; substance use; posttraumatic stress disorder

An extensive literature documents the scope and impact of intimate partner aggression victimization among medical populations, including those experiencing chronic pain (Balousek, Plane, & Fleming, 2007). Relative to this research, little work has examined rates of perpetration of partner aggression in primary care medical settings, and none has focused on patients with chronic pain. Therefore, in the current investigation, we set out to examine the prevalence of behaviors reflecting intimate partner aggression perpetration among a sample of chronic pain patients, as well as potential correlates associated with these forms of aggression.

Chronic pain has been linked to psychiatric factors that are characterized by negative affect and impulsive behavior, and that confer risk for aggression. In particular, patients with chronic pain are likely to evidence heightened posttraumatic stress disorder (PTSD),

DOI: 10.1891/0886-6708.25.5.649

depression, and substance use problems (Larson et al., 2007; Liebschutz et al., 2007; McWilliams, Cox, & Enns, 2003), all of which are strongly linked to relationship aggression perpetration in other populations (Jordan et al., 1992; Stuart, Moore, Gordon, Ramsey, & Kahler, 2006; Taft et al., 2005). Seminal theories of aggression, such as Berkowitz' cognitive-neoassociationistic model (Berkowitz, 1990), hold that those who experience more frequent and severe negative affect also experience heightened feelings, thoughts, and memories related to anger and have a higher propensity for aggressive behavior. Problematic substance use is further likely to decrease positive communication behaviors and disinhibit aggressive behavior (Leonard & Roberts, 1998), particularly in the presence of anger and heightened negative affect (Eckhardt, 2007).

While no theoretical or empirical models have been developed to explain the possible association of chronic pain with relationship aggression, Fishbain and colleagues (Fishbain, Cutler, Rosomoff, & Steele-Rosomoff, 2000) have developed a clinical model of patient violence toward physicians that describes some potential explanatory mechanisms. Specifically, this model, which has received some recent empirical support (Bruns, Disorbio, & Hanks, 2007), highlights the role of problematic and stressful interpersonal relationships with those involved in the patients' care, negative affect, physical symptom factors such as level of pain and perceptions of functional health and disability, and potential substance abuse. Analogous processes are likely to occur within the context of an intimate relationship, which has its own set of stressors, such as pain-related occupational and family role changes, financial difficulties, and impaired sexual functioning (Schwartz, Slater, & Birchler, 1996). Such relationship strains are likely to independently or jointly increase risk for aggression along with negative affect and possible substance abuse problems that accompany chronic pain.

We expected that variables reflecting negative affect and behavioral disinhibition would distinguish primary care chronic pain patients who report intimate partner aggression perpetration from their nonaggressive counterparts. Specifically, we examined PTSD, depression, and substance use disorders (SUD) as potential correlates of aggression. Consistent with the Fishbain model (Fishbain et al., 2000), it was also hypothesized that chronic pain severity and indices of physical and mental health disability would be associated with higher relationship aggression. Demographic and background correlates (age, gender, and race/ethnicity) were also explored, and we considered the role of victimization experience since much intimate aggression is bidirectional in community-based samples (Johnson & Ferraro, 2000) and individuals may aggress out of self-defense (White, Smith, Koss, & Figueredo, 2000).

METHODS

Participants

Participants were 597 patients who were 18 to 60 years of age, spoke English, endorsed pain of 3 months or more, reported use of any analgesic medication (over-the-counter or prescription) in the prior month, and had a scheduled primary care appointment. Of the 825 who met eligibility criteria for the study, 597 (76%) agreed to participate. When comparing screening questions responses between those who enrolled and those who declined, enrollees were more likely to be African American (61% vs. 55%, $p < .05$), less likely to take over-the-counter pain medication (67% vs. 79%, $p < .001$), and more likely to take

opioid pain medication (41% vs. 30%, $p < .01$). Age and gender were not different. Overall, the sample averaged 45.8 years of age and was 58.6% female, 60.8% African American, 27.8% with less than a high school education, 61.1% with a reported income less than US$20,000, 60.5% unemployed, and the majority experienced high pain limitation.

Trained research interviewers consecutively approached patients in primary care waiting rooms of an academic, urban, safety-net hospital primary care practice. Potential participants were asked to complete a written screening instrument about their pain, analgesia use, and demographic characteristics. Written informed consent was obtained from eligible and interested patients. All study measures were administered via interviews that lasted 45 to 90 min and participants were compensated US$10. Recruitment occurred between February 2005 and August 2006. The Boston University Medical Center Institutional Review Board approved the study, and National Institutes of Health issued a Certificate of Confidentiality.

Measures

Aggression perpetration was assessed with three questions taken from Wave III of the Add Health Home Questionnaire (Carolina Population Center, n.d.; Fang & Corso, 2007). Each question represented separate dependent variables: (1) Low-level aggression: Have *you* ever threatened your partner with violence, pushed or shoved [him or her], or thrown something at [him or her] that could hurt? (2) Injury: Has your partner ever had an injury, such as a sprain, bruise, or cut because of a fight with you? and (3) Sexual coercion: Have *you* ever insisted on or made your partner have sexual relations with you when [he or she] didn't want to? Participants reported on each outcome using a yes/no dichotomous scale. After each positive response, participants were asked the year of the last perpetration behavior. Each perpetrating behavior was analyzed as a separate outcome. Partner aggression perpetration assessed using the Add Health Questionnaire has been shown to be significantly associated with an index of general aggression perpetration in young adulthood, attesting to the construct validity of this outcome measure (Herrera, Wiersma, & Cleveland, 2008). Intimate partner victimization was measured using the same three questions. Any of the three victimization experiences constituted victimization in bivariate and regression analyses.

The Composite International Diagnostic Interview (CIDI; World Health Organization, 1997) was used to measure PTSD ever (lifetime) or in the past year (current). The CIDI has been shown to have good test-retest and interrater reliability and good validity (Andrews & Peters, 1998; Wittchen, 1994).

Major depression was measured using the Patient Health Questionnaire (PHQ) for Depression (Kroenke, 2002). The PHQ is a nine-item measure examining past 2 week major depression with items rated on a 4-point scale and total scores ranging from 9 to 27. The psychometric properties of the measure have been previously demonstrated (Kroenke, 2002).

SUD was defined as meeting *DSM-IV* criteria for any drug abuse or dependence ever, and/or past year alcohol dependence as measured by the CIDI version 2.1 for drug disorder (World Health Organization, 1997) and Short-Form (SF) for alcohol dependence (World Health Organization, 1997). Past year SUD included active diagnosis in the past 12 months.

Pain-related disability (limiting or nonlimiting) was measured using the Graded Chronic Pain Scale, a seven-item validated measure of pain and disability that includes

two subscales: Chronic Pain Intensity and Disability Points (Von Korff, Ormel, Keefe, & Dworkin, 1992). Scoring involves categorizing the participant into one of five pain grades: pain free, low disability-low intensity, low disability-high intensity, high disability-moderately limiting, and high disability-severely limiting.

Health-related quality of life was measured with the SF-12 Mental Health and SF-12 Physical Health composite scores (Ware, Kosinski, & Keller, 1996). This measure is derived from the SF-36 Health Survey and is scored using norm-based scoring. Several studies in both medical and general populations have shown the SF-12 to have good reliability and validity (Gandek et al., 1998; Salyers, Bosworth, Swanson, Lamb-Pagone, & Osher, 2000; Ware et al., 1996).

Analysis

This is a secondary analysis of a cross-sectional study of primary care patients with chronic pain designed to look at correlates of pain, SUD, and violence-related mental health problems. After computing descriptive statistics for the aggression outcomes, bivariate analyses were performed examining differences in characteristics associated with each perpetrating behavior. Logistic regression models were created using those variables found to be significantly associated with aggression perpetration at the bivariate level, as well as victimization for all models.

RESULTS

Descriptive Statistics for Aggression

Descriptive statistics for the study correlates are reported in Table 1. As is shown in Table 2, almost one-third of participants (30%) reported perpetrating low-level aggression toward their partner, and less than half of the sample (44%) reported low-level aggression victimization. The prevalence of injury stemming from intimate partner physical aggression victimization (33%) was approximately 3 times greater than was the prevalence of participants reporting that they injured their partner (12%). Five percent of participants reported engaging in sexual coercion, and 20% of participants indicated that their partners sexually coerced them. Participants reported a mean of about 10 years since the last perpetration behavior (9.9 for low-level aggression and sexual coercion and 13.2 for injuring partner) and 12 years since last victim experience (11.2 for low-level aggression, 12.1 for injury by partner, and 11.8 for sexual coercion).

We further examined intimate partner aggression prevalence by victim–perpetrator status (victim-only status, perpetrator-only status, or both victim and perpetrator). As Table 3 indicates, most participants who reported aggression perpetration also reported victimization (of any type). For example, 85% of participants who reported low-level intimate partner aggression perpetration also reported victimization. It is not known whether this was bidirectional in the same relationship or victimization and perpetration in different relationships.

Correlates of Intimate Partner Aggression Perpetration

Several potential correlates were examined as factors that may distinguish those who report intimate partner aggression perpetration versus those who do not. Results from these analyses are presented in Table 4. A gender effect was found, such that women were more

TABLE 1. Descriptive Statistics for Study Correlates (*N* = 597)

Variable	*N*	%
Female	350	58.6
Race		
Black	363	60.8
Hispanic	59	9.9
White	103	17.3
Other	70	11.7
Victimization	316	52.9
Current PTSD	123	20.6
Lifetime PTSD	219	36.7
Depression	249	41.7
Current or lifetime SUD	256	42.9
Limiting pain	535	89.6
	M	*SD*
Age in years	45.8	9.6
SF-12 physical health	36.5	11.7
SF-12 mental health	42.2	12.7

Note. Victimization includes at least one of the three forms of aggression (i.e., low-level aggression, injury, sexual coercion). PTSD = Posttraumatic Stress Disorder; SUD = Substance Use Disorder; SF-12 = Short Form-12 Physical and Mental Health Related Quality of Life.

likely to report perpetration of low-level aggression, and men were more likely to report sexual coercion of a partner. Partner aggression victimization was strongly associated with both low-level aggression and partner injury, and its association with sexual coercion approached significance. Lifetime PTSD represented a significant correlate of low-level aggression and partner injury, whereas a current diagnosis of PTSD was associated only with low-level aggression. Current major depression, on the other hand, was not significantly associated with any form of aggression, though its association with low-level aggression approached significance. SUD represented a significant correlate for all three outcomes. Mental health–related quality of life score was lower (worse) in perpetrators of low-level aggression but not the other types of aggression.

Regression Analyses

Table 5 reports the outcomes of regression models predicting the three outcomes. In Model 1, female gender, any victimization, lifetime PTSD, and SUD were associated with low-level aggression perpetration, whereas mental health–related quality of life was not. When we substituted current PTSD for lifetime PTSD, it was not statistically significant (data not

TABLE 2. Intimate Partner Aggression Descriptives (*N* = 597)

	N	%
Perpetration		
Low-level aggression	180	30.15
Injury	74	12.40
Sexual coercion	30	5.03
No perpetrator experiences	382	65.64
Any 1 perpetrator experience	124	20.77
Any 2 perpetrator experiences	67	11.22
Any 3 perpetrator experiences	9	1.51
	M	*SD*
Mean number of perpetration behaviors	0.49	0.76
	N	%
Victimization		
Low-level aggression	262	43.90
Injury	199	33.34
Sexual coercion	119	19.93
No victim experiences	281	48.28
Any 1 victim experience	96	16.08
Any 2 victim experiences	125	20.93
All 3 victim experiences	80	13.40
	M	*SD*
Mean number of victimization experiences	1.01	1.12

shown). For Model 2, any victimization and SUD were associated with higher infliction of injury, whereas lifetime PTSD was not. For Model 3, female gender was associated with less sexual coercion, while any victimization experience was associated with more sexual coercion and SUD was not associated with this outcome.

DISCUSSION

High rates of intimate partner aggression perpetration and victimization were reported in this sample of primary care patients with chronic pain recruited from an urban academic practice, with almost one-third reporting perpetration of low-level aggression and almost one half of the sample reporting low-level aggression victimization. More than 12% of the

TABLE 3. Intimate Partner Aggression Behavior by Victim–Perpetrator Status

	Victim		Perpetrator		Both	
	N	%	*N*	%	*N*	%
Perpetration						
Low-level aggression			27	15.1	152	84.9
Injury			9	12.0	66	88.0
Sexual coercion			10	32.3	21	67.7
	M	*SD*	*M*	*SD*	*M*	*SD*
Mean number of perpetration behaviors			1.31	0.63	1.45	0.57
	N	%	*N*	%	*N*	%
Victimization						
Low-level aggression	116	43.9			148	56.1
Injury	81	40.5			119	59.5
Sexual coercion	50	40.9			72	59.0
	M	*SD*	*M*	*SD*	*M*	*SD*
Mean number of victimization experiences	1.82	0.76			2.05	0.75

sample reported the infliction of injuries on their partner, and rates of injury victimization were almost 3 times higher. Approximately 5% of this sample reported engaging in sexual coercion, while rates of sexual coercion victimization were 4 times higher. Considering data on relationship aggression rates obtained from representative sample studies of the general population (Coker et al., 2002), and being mindful of the use of different aggression measures across studies, current findings suggest elevated rates of aggression occurring in the intimate relationships of patients experiencing chronic pain.

Reports of higher rates of intimate partner aggression victimization than perpetration are consistent with the focus of the broader literature that has emphasized associations between abuse victimization experiences and chronic pain (Bailey, Freedenfeld, Kiser, & Gatchel, 2003; Balousek et al., 2007; Walsh, Jamieson, Macmillan, & Boyle, 2007). It is important to note, however, that individuals tend to underreport their intimate relationship perpetration behavior relative to their victimization due to social desirability and other biases (Moffitt et al., 1997). Thus, perpetration reports in this study are likely to represent underestimates, and the true rates of aggression victimization and perpetration are likely to be more comparable than current study findings indicate. In addition, study findings indicate that a number of correlates were associated with intimate relationship aggression perpetration in this sample, even when controlling for victimization experiences, suggesting that aggression perpetrated in this sample was not exclusively due to acts of self-defense or bidirectional aggression.

TABLE 4. Bivariate Correlates of Perpetration Behaviors

	Low-Level Aggression Perpetration					Injury Perpetration					Sexual Coercion Perpetration				
	Yes (n = 179)		No (n = 403)			Yes (n = 75)		No (n = 507)			Yes (n = 31)		No (n = 550)		
Variable	n	%	n	%	p Value	n	%	n	%	p Value	n	%	n	%	p Value
Gender															
Male	56	31.4	185	45.9	.001	26	34.7	215	42.4	.20	23	74.2	218	39.6	.0001
Female	123	68.7	218	54.1		49	65.3	292	57.6		8	25.8	332	60.4	
Race															
Black	108	60.3	246	61.4	.96	45	60.0	309	61.2	.94	20	64.5	333	60.8	.80
White	32	17.9	68	16.9		14	18.7	86	17.0		4	12.9	96	17.5	
Other	39	21.8	87	21.7		16	21.3	110	21.8		7	22.6	119	21.7	
Victimization															
Yes	152	84.9	149	36.9	<.0001	66	88.0	235	46.4	<.0001	21	67.7	280	50.9	.07
No	27	15.1	254	63.0		9	12.0	272	53.6		10	32.3	270	49.1	
Current PTSD															
Yes	55	30.8	67	16.6	.0001	17	22.7	105	20.7	.69	10	32.3	112	20.4	.11
No	124	69.3	336	83.4		58	77.3	402	79.3		21	67.7	438	79.6	
Lifetime PTSD															
Yes	100	55.9	116	28.9	<.0001	49	52.0	177	34.9	.004	12	38.7	204	37.1	.86

No	79	44.1	287	71.2		36	48.0	330	65.1		19	61.3	346	62.9	
Major depression															
Yes	74	41.3	134	33.3	.06	30	40.0	178	35.1	.41	11	35.5	197	35.5	.97
No	105	58.7	269	66.8		45	60.0	329	64.9		20	64.5	353	64.2	
Substance use disorder															
Past year	55	30.7	64	15.9	<.0001	22	29.3	97	19.1	<.0001	12	38.7	107	19.5	.02
Prior to past year	47	26.3	80	19.9		28	37.3	99	29.5		8	25.8	119	21.6	
No lifetime SUD	77	43.0	259	64.3		25	33.3	311	61.3		11	35.5	324	58.9	
Limiting pain															
Yes	164	91.6	358	88.8	.31	67	89.3	455	89.7	.91	27	87.1	494	99.8	.55
No	15	8.4	45	11.2		8	10.7	52	10.3		4	12.9	56	10.2	
	M	*SD*	*M*	*SD*	*p* Value	*M*	*SD*	*M*	*SD*	*p* Value	*M*	*SD*	*M*	*SD*	*p* Value
Age in years	46.43	8.93	45.51	9.94	.29	46.56	9.1	45.7	9.7	.5	45.1	10.2	45.8	9.6	.87
SF-12 physical health score	36.17	11.22	36.70	11.80	.6	35.74	12.2	36.7	11.5	.5	37.3	11.6	36.5	11.4	.69
SF-12 mental health score	40.28	13.14	42.84	12.36	.02	40.49	12.6	42.3	12.9	.3	42.92	11.5	41.9	12.7	.69

Note. Victimization includes at least one of the three forms of aggression (i.e., low-level aggression, injury, sexual coercion). PTSD = Posttraumatic Stress Disorder; SUD = Substance Use Disorder; SF-12 = Short Form-12 Physical and Mental Health–Related Quality of Life.

TABLE 5. Characteristics Associated With Intimate Partner Aggression Perpetration

Model	Odds Ratio (95% CI)
1. Low-level aggression	
Female vs. Male	1.97 (1.25–3.11)
Any victimization—Yes vs. No	7.18 (4.45–11.59)
Lifetime PTSD—Yes vs. No	1.81 (1.18–2.77)
Any SUD—Yes vs. No	2.23 (1.43–3.47)
SF-12 mental health score	1.01 (0.99–1.02)
2. Injury	
Any victimization—Yes vs. No	7.12 (3.40–14.90)
Lifetime PTSD—Yes vs. No	1.12 (0.66–1.89)
Current/lifetime SUD—Yes vs. No	2.42 (1.42–4.13)
3. Sexual coercion	
Female vs. Male	0.22 (0.09–0.53)
Any victimization—Yes vs. No	2.34 (1.04–5.30)
Current/lifetime SUD—Yes vs. No	1.54 (0.68–3.47)

Note. Victimization includes at least one of the three forms of aggression (i.e., low-level aggression, injury, sexual coercion). PTSD = Posttraumatic Stress Disorder; SUD = Substance Use Disorder; SF-12 = Short Form-12 Physical and Mental Health–Related Quality of Life.

Consistent with the Fishbain model (Fishbain et al., 2000) adapted for intimate partner aggression, it was predicted that variables reflecting negative affect and behavioral disinhibition, as well as chronic pain severity and disability would emerge as significant correlates of aggression perpetration. Among these predictors, SUD generally emerged as the strongest relative predictor. This correlate was associated with each of the three aggression perpetration outcomes at the bivariate level and both measures reflecting nonsexual aggression when statistically accounting for the other significant correlates. Problematic substance use leads to disinhibition of aggressive impulses (Eckhardt, 2007; Leonard & Roberts, 1998), and previous research indicates that substance abuse is associated with violent ideation in this population (Bruns et al., 2007). Substance use may be particularly problematic in the context of PTSD and poor mental health functioning, which were also associated with aggression perpetration at the bivariate level. PTSD and not poor mental health functioning was associated with low-level aggression perpetration in the context of the other significant correlates.

Women appeared to report more low-level aggression than men, while men reported more engagement in sexual coercion behavior. These findings are generally consistent with the broader literature on intimate partner aggression perpetration. A meta-analysis by Archer (2000) indicated that women engage in slightly higher rates of

noninjurious intimate aggression than men, particularly in community-based samples (Archer, 2000). Men's aggression is more likely to lead to victim injury, though current study findings did not find such gender differences. Regarding differences in sexual coercion, previous research indicates that men engage in higher levels of sexual coercion or sexual aggression than women (Hartwick, Desmarais, & Hennig, 2007; Stets & Pirog-Good, 1987).

The current investigation has some important clinical implications. Intimate partner aggression victimization as well as perpetration appears to be heightened in the chronic pain population, suggesting that increased screening, prevention, and intervention efforts focused on partner aggression are warranted for these individuals. Such efforts should target both men and women, as current study findings suggest that although some gender differences were noted, both genders may engage in or experience intimate partner aggression. It appears that interventions that target SUD in particular, as well as symptoms of PTSD, may be especially effective in reducing aggression. Couples-based interventions also appear warranted for this population, as the aggression reported in this study suggests that it may frequently be bidirectional in nature, and victimization was a robust predictor of perpetration. Previous work indicates that the response of the intimate partner to a patient's negative pain behaviors can serve as powerful determinants of adjustment and the maintenance of such behaviors, lending further support for couples-based intervention approaches (Burns, Johnson, Mahoney, Devine, & Pawl, 1996; Cano, Gillis, Heinz, Geisser, & Foran, 2004; Cano & Leonard, 2006; Newton-John & Williams, 2006; Romano et al., 1992; Schwartz et al., 1996). However, couples therapy may be contraindicated in cases of moderate- to severe-aggression or in the presence of a pattern of coercive control in the relationship.

The cross-sectional nature of this study precludes us from drawing firm conclusions regarding the directionality of obtained associations. Findings that much of the aggression may have occurred several years prior to study participation (taking into account the previously described possible deflated self-reported rates of aggression) suggest that aggression victimization led to the experience of chronic pain in this sample. Moreover, aggression perpetration may also lead to higher levels of chronic pain because anger expression may alienate patients from their partners and other sources of support (Burns et al., 1996), and several other psychological, biological, and genetic mechanisms have been proposed for this relationship (Bruehl, Chung, & Burns, 2006). Prospective designs are needed to more fully examine the directionality of associations among the variables investigated in the current study. It is perhaps most likely that associations among chronic pain, aggression victimization and perpetration, and the correlates of interest are bidirectional in nature. Future research in this area should also utilize more comprehensive measures of different forms of physical, psychological, and sexual intimate partner aggression and should obtain reports from both members of the couple. Finally, sampling was limited to one primary care setting in one locale. It is possible that findings would not generalize to other settings or study sites.

Despite these limitations, this study represents an initial attempt to examine reports of intimate relationship aggression perpetration in a sample of chronic pain patients, including correlates of such aggression. Findings suggest relatively high rates of aggression perpetration and victimization in this sample and highlight the role of substance use problems in particular as a correlate of perpetration. Additional work is needed to better understand the nature and scope of the relationship aggression problem in patients experiencing chronic pain and to ultimately reduce aggression and enhance intimate relationships in this population.

REFERENCES

Andrews, G., & Peters, L. (1998). The psychometric properties of the composite international diagnostic interview. *Social Psychiatry and Psychiatric Epidemiology, 33*(2), 80–88.

Archer, J. (2000). Sex differences in aggression between heterosexual partners: A meta-analytic review. *Psychological Bulletin, 126*, 651–680.

Bailey, B. E., Freedenfeld, R. N., Kiser, R. S., & Gatchel, R. J. (2003). Lifetime physical and sexual abuse in chronic pain patients: Psychosocial correlates and treatment outcomes. *Disability and Rehabilitation, 25*, 331–342.

Balousek, S., Plane, M. B., & Fleming, M. (2007). Prevalence of interpersonal abuse in primary care patients prescribed opioids for chronic pain. *Journal of General Internal Medicine, 22*, 1268–1273.

Berkowitz, L. (1990). On the formation and regulation of anger and aggression: A cognitive-neoassociationistic analysis. *American Psychologist, 45*, 494–503.

Bruehl, S., Chung, O. Y., & Burns, J. W. (2006). Anger expression and pain: An overview of findings and possible mechanisms. *Journal of Behavioral Medicine, 29*, 593–606.

Bruns, D., Disorbio, J. M., & Hanks, R. (2007). Chronic pain and violent ideation: Testing a model of patient violence. *Pain Medicine, 8*, 207–215.

Burns, J. W., Johnson, B. J., Mahoney, N., Devine, J., & Pawl, R. (1996). Anger management style, hostility and spouse responses: Gender differences in predictors of adjustment among chronic pain patients. *Pain, 64*, 445–453.

Cano, A., Gillis, M., Heinz, W., Geisser, M., & Foran, H. (2004). Marital functioning, chronic pain, and psychological distress. *Pain, 107*, 99–106.

Cano, A., & Leonard, M. (2006). Integrative behavioral couple therapy for chronic pain: Promoting behavior change and emotional acceptance. *Journal of Clinical Psychology, 62*, 1409–1418.

Carolina Population Center. (n.d.). *Add Health Questionnaire: Wave III—In Home Questionnaire Code Book* (The National Longitudinal Study of Adolescent Health). Chapel Hill: University of North Carolina.

Coker, A. L., Davis, K. E., Arias, I., Desai, S., Sanderson, M., Brandt, H. M., et al. (2002). Physical and mental health effects of intimate partner violence for men and women. *American Journal of Preventive Medicine, 23*, 260–268.

Eckhardt, C. I. (2007). Effects of alcohol intoxication on anger experience and expression among partner assaultive men. *Journal of Consulting and Clinical Psychology, 75*(1), 61–71.

Fang, X., & Corso, P. S. (2007). Child maltreatment, youth violence, and intimate partner violence: Developmental relationships. *American Journal of Preventive Medicine, 33*, 281–290.

Fishbain, D. A., Cutler, R. B., Rosomoff, H. L., & Steele-Rosomoff, R. (2000). Risk for violent behavior in patients with chronic pain: Evaluation and management in the pain facility setting. *Pain Medicine, 1*(2), 140–155.

Gandek, B., Ware, J. E., Aaronson, N. K., Apolone, G., Bjorner, J. B., Brazier, J. E., et al. (1998). Cross-validation of item selection and scoring for the SF-12 Health Survey in nine countries: Results from the IQOLA Project. International Quality of Life Assessment. *Journal of Clinical Epidemiology, 51*, 1171–1178.

Hartwick, C., Desmarais, S., & Hennig, K. (2007). Characteristics of male and female victims of sexual coercion. *Canadian Journal of Human Sexuality, 16*, 31–44.

Herrera, V., Wiersma, J., & Cleveland, H. (2008). The influence of individual and partner characteristics on the perpetration of intimate partner violence in young adult relationships. *Journal of Youth and Adolescence, 37*, 284–296.

Johnson, M., & Ferraro, K. (2000). Research on domestic violence in the 1990s: Making distinctions. *Journal of Marriage and the Family, 62*, 948–963.

Jordan, B. K., Marmar, C. R., Fairbank, J. A., Schlenger, W. E., Kulka, R. A., Hough, R. L., et al. (1992). Problems in families of male Vietnam veterans with posttraumatic stress disorder. *Journal of Consulting and Clinical Psychology, 60*, 916–926.

Kroenke, K. (2002). The PHQ-9: A new depression and diagnostic severity measure. *Psychiatric Annals, 32*, 509–521.

Larson, M. J., Paasche-Orlow, M., Cheng, D. M., Lloyd-Travaglini, C., Saitz, R., & Samet, J. H. (2007). Persistent pain is associated with substance use after detoxification: A prospective cohort analysis. *Addiction, 102*, 752–760.

Leonard, K. E., & Roberts, L. J. (1998). The effects of alcohol on the marital interactions of aggressive and nonaggressive husbands and their wives. *Journal of Abnormal Psychology, 107*, 602–615.

Liebschutz, J., Saitz, R., Brower, V., Keane, T. M., Lloyd-Travaglini, C., Averbuch, T., et al. (2007). PTSD in urban primary care: High prevalence and low physician recognition. *Journal of General Internal Medicine, 22*, 719–726.

McWilliams, L. A., Cox, B. J., & Enns, M. W. (2003). Mood and anxiety disorders associated with chronic pain: An examination in a nationally representative sample. *Pain, 106*, 127–133.

Moffitt, T., Caspi, A., Krueger, R. M. L, Margolin, G., Silva, P., & Sydney, R. (1997). Do partners agree about abuse in their relationship? A psychometric evaluation of interpartner aggression. *Psychological Assessment, 47*, 47–56.

Newton-John, T. R., & Williams, A. C. (2006). Chronic pain couples: Perceived marital interactions and pain behaviours. *Pain, 123*, 53–63.

Romano, J. M., Turner, J. A., Friedman, L. S., Bulcroft, R. A., Jensen, M. P., Hops, H., et al. (1992). Sequential analysis of chronic pain behaviors and spouse responses. *Journal of Consulting and Clinical Psychology, 60*, 777–782.

Salyers, M. P., Bosworth, H. B., Swanson, J. W., Lamb-Pagone, J., & Osher, F. C. (2000). Reliability and validity of the SF-12 health survey among people with severe mental illness. *Medical Care, 38*, 1141–1150.

Schwartz, L., Slater, M. A., & Birchler, G. R. (1996). The role of pain behaviors in the modulation of marital conflict in chronic pain couples. *Pain, 65*, 227–233.

Stets, J., & Pirog-Good, M. (1987). Violence in dating relationships. *Social Psychology Quarterly, 50*, 237–246.

Stuart, G. L., Moore, T. M., Gordon, K. C., Ramsey, S. E., & Kahler, C. W. (2006). Psychopathology in women arrested for domestic violence. *Journal of Interpersonal Violence, 21*, 376–389.

Taft, C. T., Pless, A. P., Stalans, L. J., Koenen, K. C., King, L. A., & King, D. W. (2005). Risk factors for partner violence among a national sample of combat veterans. *Journal of Consulting and Clinical Psychology, 73*, 151–159.

Von Korff, M., Ormel, J., Keefe, F. J., & Dworkin, S. F. (1992). Grading the severity of chronic pain. *Pain, 50*, 133–149.

Walsh, C. A., Jamieson, E., Macmillan, H., & Boyle, M. (2007). Child abuse and chronic pain in a community survey of women. *Journal of Interpersonal Violence, 22*, 1536–1554.

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.

White, J. W., Smith, P. H., Koss, M. P., & Figueredo, A. J. (2000). Intimate partner aggression—what have we learned? Comment on Archer (2000). *Psychological Bulletin, 126*, 690–696.

Wittchen, H. U. (1994). Reliability and validity studies of the WHO-Composite International Diagnostic Interview (CIDI): A critical review. *Journal of Psychiatric Research, 28*(1), 57–84.

World Health Organization. (1997). *Composite International Diagnostic Interview (CIDI): Version 2.1.* Geneva, Switzerland: Author.

Acknowledgments. The authors thank Denis Rybin, MSc, for his assistance on data analysis. Funding for this study was provided by K23 DA016665 from the National Institute on Drug Abuse and K24 AA015674 from the National Institute on Alcohol Abuse and Alcoholism.

Correspondence regarding this article should be directed to Casey Taft, PhD, VA Boston Healthcare System (116B-4), 150 South Huntington Avenue, Boston, MA 02130. E-mail: casey.taft@va.gov

J Epidemiol 2010;20(4):319-328
doi:10.2188/jea.JE20090096

Original Article

Development and Validation of a 15-Item Japanese Health Knowledge Test

Yasuharu Tokuda[1], Tomoya Okubo[2], Haruo Yanai[3], Nobutaka Doba[4], and Michael K. Paasche-Orlow[5]

[1]Institute of Clinical Medicine, Graduate School of Comprehensive Human Sciences, University of Tsukuba, Tsukuba, Ibaraki, Japan
[2]The National Center for University Entrance Examinations, Tokyo, Japan
[3]St Luke's Graduate School of Nursing, Tokyo, Japan
[4]The Life Planning Center, Tokyo, Japan
[5]Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, Boston, USA

Received June 22, 2009; accepted March 1, 2010; released online June 12, 2010

ABSTRACT

Background: Health literacy affects the acquisition of health knowledge and is thus linked to health outcomes. However, few scales have been developed to assess the level of health knowledge among the general public.

Methods: The 15-item Japanese Health Knowledge Test (J-HKT) was developed by using item response theory to score an item pool. We examined the construct validity of the J-HKT in relation to health literacy items, and analyzed the sociodemographic and behavioral factors associated with poor health knowledge.

Results: We enrolled 1040 adult participants (mean age, 57 years; women, 52%). The 15 items that best identified people with poor health knowledge were selected. For all items on the J-HKT, the information function curves had a peak in the negative spectrum of the latent trait. As compared with participants reporting high levels of income, educational attainment, and literacy, those with low levels of income, education, and literacy had a lower total score on the J-HKT. As compared with non/light drinkers, moderate and heavy drinkers had lower total scores on the J-HKT.

Conclusions: The J-HKT may prove useful in measuring health knowledge among the general public, and in identifying and characterizing those with poor health knowledge.

Key words: health knowledge; health literacy; socioeconomic status

INTRODUCTION

A growing body of evidence supports the impact of low health literacy on the health of individuals[1]; therefore, recent attention has focused on the elucidation of potential causal pathways linking low health literacy to poor health.[2,3] Among the mechanisms that mediate the influence of health literacy on the health of individuals, the effect of health literacy on health knowledge may be one of the most consistent and critical factors.[4–6] It has been proposed that low health literacy leads to poor health knowledge and, ultimately, to worse health outcomes, because people with low health literacy have difficulty in acquiring the health knowledge necessary to navigate the healthcare system and to practice effective self-care.

Patients with poor knowledge of illness prevention and chronic diseases have lower adherence to medical instructions and are more likely to have high-risk health behaviors.[7–10] Thus, these individuals are less likely to utilize healthcare services, such as recommended vaccination and health screening programs.[11–13] In addition, during both acute and chronic illnesses, the quality of self-care is poor among those with limited knowledge, which may manifest in the greater use of potentially harmful complementary or alternative medicine.[14]

Many studies have evaluated the relationship between health literacy and health knowledge.[4,5,7,8,15,16] These have mostly focused on patients with specific illnesses, such as asthma, diabetes, congestive heart failure, hypertension, and human immunodeficiency virus infection; few have evaluated the association between low health literacy and poor health knowledge in the general public. In patients with chronic diseases, the relationship between health literacy and health knowledge of a particular disease has been confirmed.[4,5,7,8,15,16] In order to better understand the relationship between health literacy and health knowledge, and to help target education and guide disease prevention for the general public, it would be useful to examine the

Address for correspondence. Prof. Yasuharu Tokuda, 3-2-8 Miya-machi, Ibaraki, Mito City 3100015, Japan (e-mail: tokuyasu@orange.ocn.ne.jp).

relationship between health knowledge and health literacy in the general public. However, this objective cannot be realized without a tool to assess general health knowledge. Such a tool would be particularly useful if it were short, if it could differentiate among people at the low end of the health knowledge spectrum, and if it could be administered in a mode other than in-person interview. Thus, in the present study, we used nominal categories modeling of item response theory (IRT) analysis to develop a test of general health knowledge for Japanese adults. To evaluate construct validity, we examined the association between this health knowledge test and health literacy. In addition, we identified the sociodemographic and health behavioral factors that were significantly associated with poor health knowledge.

METHODS

Study participants

The data for this study were collected from responses to a national cross-sectional online survey conducted from 3 July to 8 July 2008. Institutional review board approval was obtained from the National Institute of Japanese Language. Japan was divided into 10 regions: Hokkaido, Tohoku, Kanto, Tokai, Keihin, Hokuriku, Kyouhanshin, Chugoku, Shikoku, and Kyushu. The number of potential participants was determined within each region from a panel of people registered by Yahoo JAPAN Co. (Tokyo, Japan) by means of probability sampling proportionate to age and sex, using Japanese national census data of population distributions for people aged 30–90 years in 2007. People younger than 30 years were excluded because our aims included evaluation of the potential association between health knowledge and final educational attainment. In addition, health care workers, such as physicians, nurses, hospital workers, and public health workers, were excluded. No gifts or payments were given for participating in the survey.

Data collection

The survey gathered demographic and socioeconomic data, as well as responses to the questionnaire for health literacy and the test of health knowledge. Demographic data included age, sex, annual income, education, and occupation. Regarding annual income, cutoff points of 2, 4, 6, and 8 million Japanese Yen (JY) were used to generate 5 income categories (the average exchange rate for 1 US dollar in July 2008 was about 100 JY). We used these income cutoffs because the National Tax Agency regards an income of 2 million JY as the cutoff level for low-wage workers and reports the income distribution in this fashion. For educational attainment, 5 categories were used (did not graduate high school, high school graduate, vocational school, short-term college, and undergraduate/postgraduate degree). For occupational status, 5 categorical levels were included: working full-time, homemaker, working part-time, retired, and not currently working. Survey items also assessed current and past smoking, current alcohol use, and chronic conditions (cancer, cardiovascular disease, hypertension, diabetes, arthritis, asthma or chronic obstructive pulmonary disease, and depression), as previously described.[2]

Current alcohol consumption was categorized into 3 categories: non/light, moderate, and heavy. Non/light drinkers were defined as those who drank less than once a week; moderate and heavy users included those who drank at least once a week. In addition, heavy users were defined as those who drank in a day ≥3 glasses of beer, ≥540 ml of Japanese sake (*nihonshu*), three-quarters of a bottle or more of wine, or ≥180 ml of whisky. All remaining participants were defined as moderate users.

Health literacy was measured by self-report using 2 validated screening questions.[17,18] Specifically, we asked: "How often do you have problems learning about your medical condition because of difficulty understanding written information?" (Item 1: "Problems learning") and "How often do you have someone help you read hospital materials?" (Item 2: "Help reading"). The 5-point Likert response scale was, "Never", "Occasionally", "Sometimes", "Often", or "Always". These 2 items have been shown to predict scores on commonly used English-language measures of health literacy: the Short Test of Functional Health Literacy in Adults (STOFHLA) and the Rapid Estimate of Adult Literacy in Medicine (REALM).[17,18] Due to the linguistic differences between English and Japanese, English-language instruments for measuring health literacy cannot be simply translated. Thus, we used these 2 self-report items as surrogate measures of health literacy.[17–19]

Development of the Japanese Health Knowledge Test (J-HKT)

The first phase of development included item generation by a group of experts in healthcare, literacy, linguistics, and mass media. This 25-member group included physicians, nurses, pharmacists, linguists, journalists, university researchers in communication, and representatives of patient advocate groups. Each item was developed with a single correct response among a list of 4 choices. When providing the item test to the study participants, they were advised that there was a 2-minute time limit for each item. Each item was scored as correct or incorrect.

In the second phase of development, the 48-item pool was shortened using item response theory (IRT) analysis, specifically the nominal categories model. This model was proposed by Bock[20] as an extension of IRT analysis for nominally scored items. As compared with the use of a graded categories model or a binary logistic model, the nominal categories model is more effective in examining the full spectrum of contributions for each item and the possible responses in an instrument. For this purpose, we used a sample size large enough to meet the requirements of nominal categories modeling.

In the nominal categories model, the response probability p_{ijk} that respondent i with a latent trait θ_i response to category k ($k = 1, 2, \ldots, K_j$) of item j is described as follows[21]:

$$p_{ijk} = \frac{\exp(\alpha_{jk}\theta_i + \gamma_{jk})}{\sum_{k'=1}^{K_j} \exp(\alpha_{jk'}\theta_i + \gamma_{jk'})}$$

where K_j denotes the number of the category of item j. We cannot interpret the parameters of the categories independently in the nominal categories model because the equation defined for a response probability to the category contains other parameters. Thus, in order to estimate item parameters, Okubo suggested that a restriction be imposed as follows[22]:

$$\alpha_{j1} = \gamma_{j1} = 0$$

The role of the alpha parameter is that of a slope in the linear function. A larger slope implies that the item clearly discriminates the latent trait θ_i, while a smaller slope implies low discrimination. The role of the gamma parameter is that of an intercept. A larger intercept gamma suggests that the item is difficult to solve, while a smaller intercept gamma suggests it is easy to solve.

Next, the item response category characteristic curve (IRCCC) is determined by the relative relations among parameters; thus, each parameter cannot be interpreted alone. The usual method to analyze the characteristics of items is to draw the IRCCC by using the estimated parameters. The IRCCC is a multinomial logistic regression curve whose independent variable is a factor—in this case, health knowledge.

Item information functions were then generated for each item. Item information function curves were derived from the response probabilities from the IRCCCs. The standard error of measurement curve can be calculated as the reciprocal of the square root of the item information function. Item information functions describe responses at different levels of a latent trait—health knowledge in this study. A combination of all items together was used to generate the test information function, and an item reduction procedure was performed based on the item information functions. Participants with a score that was ≥1 standard deviation lower than the mean were classified as having a low score.

Phase 3 of development sought to support the validity of the J-HKT. The face validity of the J-HKT was confirmed by the aforementioned expert panel. Next, for construct validity, we hypothesized that health literacy would be associated with improved J-HKT scores and thus the association between literacy and J-HKT scores was examined by using the nonparametric test for trend across ordered groups developed by Cuzick.[23]

Associations between sociodemographic characteristics and J-HKT scores were evaluated by a logistic regression model that included age and sex, as well as additional variables found to be significant in univariate analyses. Statistical analyses were performed using R version 2.6.6 (R Foundation for Statistical Computing) and STATA 10.0 (College Station, Texas, USA), and graphics were generated using Mathematica version 6.0 (Wolfram Research, Illinois, USA). A 2-tailed P value <0.05 was considered statistically significant.

RESULTS

Of 2500 subjects randomly selected from the online panel, 1074 participated in the study (response rate, 43.0%). Among these, after deleting data from participants working in the health care industry, data for 1040 persons were available for our analysis and were considered as the final sample. Table 1 shows the sociodemographic characteristics of all participants; 52% were women and the mean age was 57 years (range, 30–90).

Table 1. Characteristics of participants (n = 1040)

Characteristic	Meen (SD) or n, %
Age (years)	57 (15)
Sex	
Male	497, 48%
Female	543, 52%
Income (Japanese Yen)	
<2 million	92, 9%
2–3.99 million	264, 25%
4–5.99 million	290, 28%
6–7.99 million	160, 15%
8 million or more	234, 23%
Education	
<Grade 12	51, 5%
High school graduate	379, 36%
Vocational school	107, 10%
Some college	139, 13%
University or graduate degree	364, 35%
Working status	
Working full-time	445, 43%
Homemaker	273, 26%
Working part-time	91, 9%
Retired	135, 13%
Currently not working	96, 9%
Smoking	
Current	200, 19%
Former	247, 24%
Never	593, 57%
Current alcohol use	
None/light	588, 57%
Moderate	407, 39%
Heavy	45, 4%
Chronic condition	
Cancer	38, 4%
Cardiovascular disease	21, 2%
Hypertension	221, 21%
Diabetes	55, 5%
Arthritis	45, 4%
Asthma or COPD	29, 3%
Depression	33, 3%

COPD = chronic obstructive pulmonary disease.

Table 2. Estimated parameters for the 15 items of the Japanese Health Knowledge Test

Item number	Alpha (slope parameter) Category				Gamma (location parameter) Category			
	1	2	3	4	1	2	3	4
1	0.00	0.07	0.46	0.36	0.00	0.88	2.04	−0.41
2	0.00	0.62	1.56	0.91	0.00	0.77	2.99	2.06
3	0.00	−1.53	−0.30	−0.45	0.00	−2.25	0.13	0.64
4	0.00	−0.69	0.33	−0.47	0.00	1.26	2.04	0.21
5	0.00	2.70	2.19	1.49	0.00	4.65	2.58	1.93
6	0.00	1.67	0.94	2.26	0.00	3.28	0.42	0.64
7	0.00	0.63	−1.79	−2.02	0.00	0.71	−2.68	−3.15
8	0.00	0.26	0.23	1.02	0.00	−0.16	1.08	2.25
9	0.00	−1.07	−0.87	−1.04	0.00	−2.04	−2.50	−2.13
10	NA	0.00	1.53	NA	NA	0.00	3.54	NA
11	0.00	0.56	NA	0.32	0.00	2.53	NA	0.89
12	0.00	−0.74	−0.87	0.24	0.00	−1.42	−0.71	−0.17
13	0.00	−1.52	0.96	NA	0.00	−2.56	0.88	NA
14	0.00	0.81	−0.33	−0.25	0.00	1.37	−0.50	−0.06
15	0.00	−1.18	−0.21	0.69	0.00	−0.91	−0.34	1.12

Values for category 1 were set to 0 for estimating parameters.
NA, not available.

The initial item pool contained 48 items that covered knowledge of body parts, diseases, hospitals, drugs, healthcare systems, health policy, and home care. The expert panel considered these 48 items to have adequate content validity, and to represent the range of patient knowledge required to understand common medical problems. Based on the item information functions of the IRT analysis for health knowledge testing in the 1040 participants, a 15-item J-HKT was produced from the initial 48-item pool (Table 2 and Supplement).

Regarding each response to individual items of the J-HKT, all IRCCCs of the J-HKT satisfied the assumption of monotonicity, ie, scores for each item were higher among participants with a higher overall J-HKT score. For most items, a greater number of intersections of probability curves of item responses was shifted to the negative spectrum of the latent trait. Figure 1 shows the item information function for individual items of the J-HKT.

To better discriminate between people with poor health knowledge and those with intermediate or higher levels of health knowledge, 15 items with the highest information function at −0.85 (those with the lowest percentile of 20% of overall scores in all participants) of latent trait θ_i were included in the J-HKT. Thus, we chose items able to differentiate among people at the low end of health knowledge; as such, the curves for all items of the J-HKT show a peak of the information functions in the negative spectrum of the latent trait.

Figure 2 shows a histogram of total scores for the 15-item J-HKT. The mean score was 4.7 and the standard deviation was 1.6; the median score was 5.0 and the mode was 4.0. The score is normally distributed, with a skewness of −0.37 and a kurtosis of −0.38. Figures 3 and 4 show the proportions of participants with poor health knowledge, by responses to the 2 health literacy items ("Problems learning" and "Help reading"). There were statistically significant associations between responses to the health literacy items and total score on the J-HKT (ie, construct validity). Figure 5 shows the item information function curve of the 15-item J-HKT, and Figure 6 depicts the standard error curve of the item information function of the 15-item J-HKT (the standard error is the reciprocal of the item information function).

Table 3 shows the distributions of total score on the J-HKT by sociodemographic characteristics, smoking, and alcohol use. Overall, 36% of participants had a score of 0–3, which was defined as poor health knowledge (ie, more than 1 standard deviation below the mean, 4.7 − 1.6 = 3.1). Age, sex, and employment status were not associated with test scores; however, participants with low income and low educational attainment were more likely to have a lower score on the J-HKT. Although smoking status was not associated with J-HKT score, those with higher current alcohol use had a lower total score on the J-HKT.

Table 4 presents the results of the logistic regression model for poor health knowledge on the J-HKT (0–3, yes versus no) adjusted for age, sex, income, education, and current alcohol use. Compared with those with an income >8 million JY, those with income ≥2 and <4 million JY were more likely to have poor health knowledge (odds ratio [OR], 1.68; 95% confidence interval [CI], 1.08–2.62) and those with an income <2 million JY were also more likely to have poor health knowledge (1.84; 1.02–3.31). In addition, as compared with university degree holders, those who had not graduated high school were also more likely to have poor health knowledge (2.08; 1.05–4.14). Regarding current alcohol use, as compared with non/light drinkers, poor health knowledge was more likely among moderate drinkers (1.53; 1.12–2.09) and heavy drinkers (2.28; 1.16–4.47).

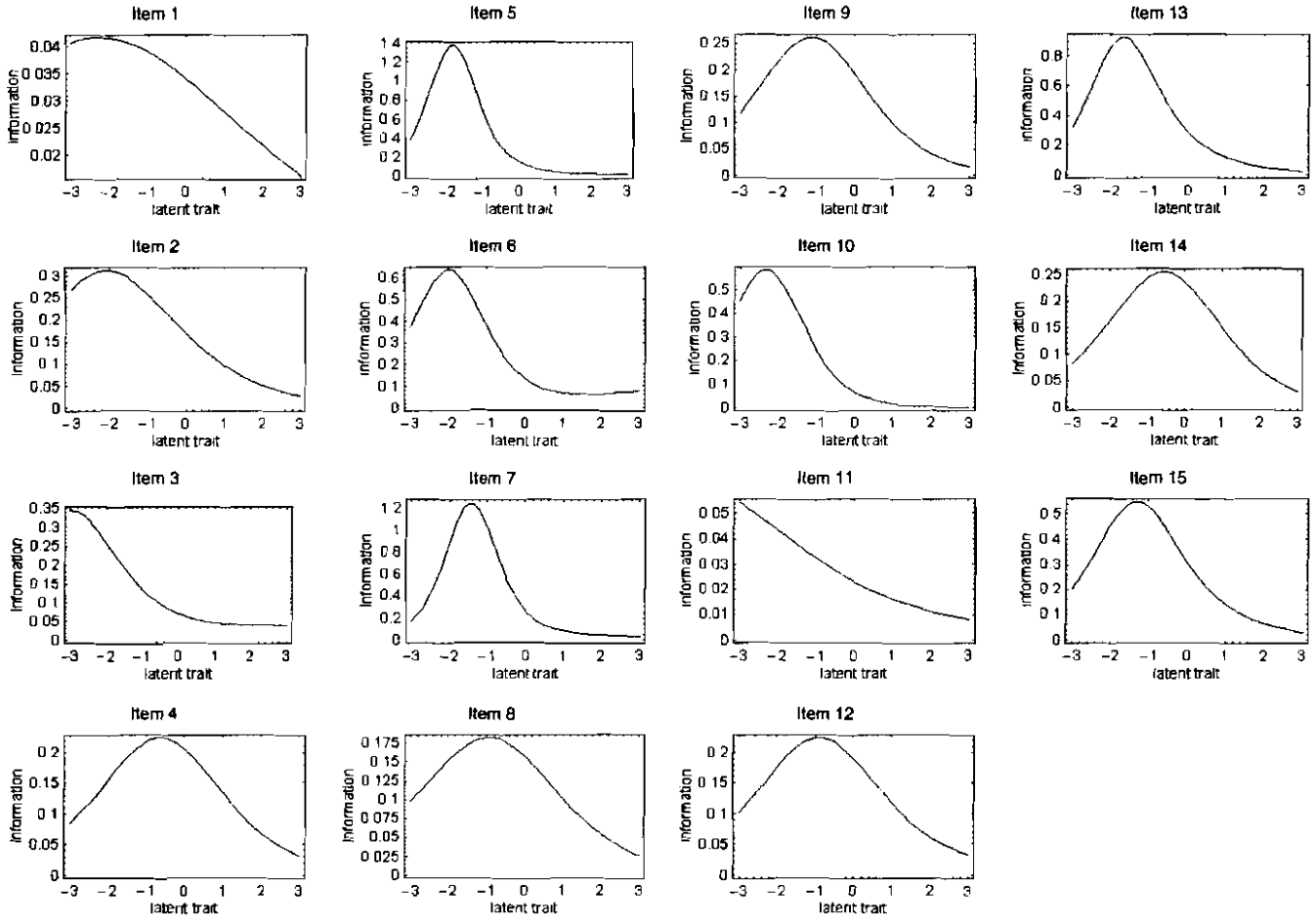


Figure 1. Item information function curves for each item of the Japanese Health Knowledge Test were generated by analysis of data from 1040 Japanese adults. The curves were derived from the response probabilities from the item response category characteristic curves. The standard error of measurement curve was calculated as the reciprocal of the square root of the item information function. Note: the scales for the y-axis differ among items.

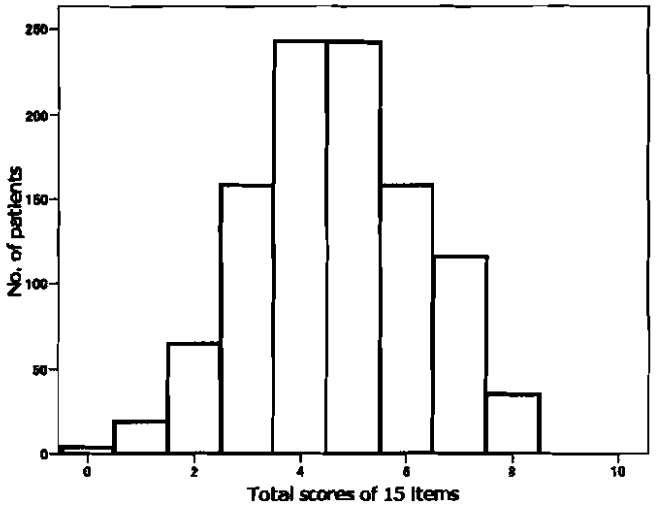


Figure 2. Histogram of total scores on the Japanese Health Knowledge Test.

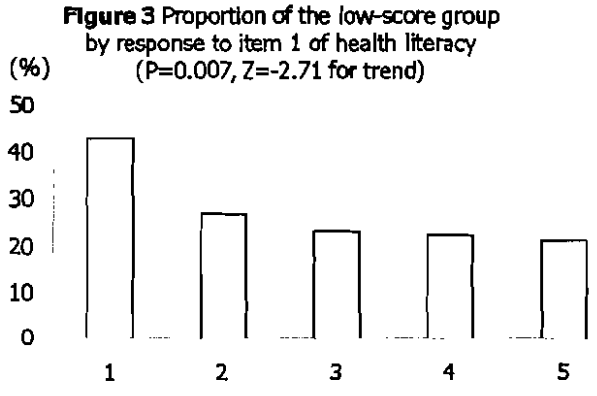


Figure 3. Proportion of participants with poor health knowledge, by response to item 1 ("Problems learning") on the health literacy test. The question was, "How often do you have problems learning about your medical condition because of difficulty understanding written information?" The 5-point Likert response scale was, "Always" (1), "Often" (2), "Sometimes" (3), "Occasionally" (4), and "Never" (5). Participants with lower literacy represented a higher proportion of those with a low score on the Japanese Health Knowledge Test.

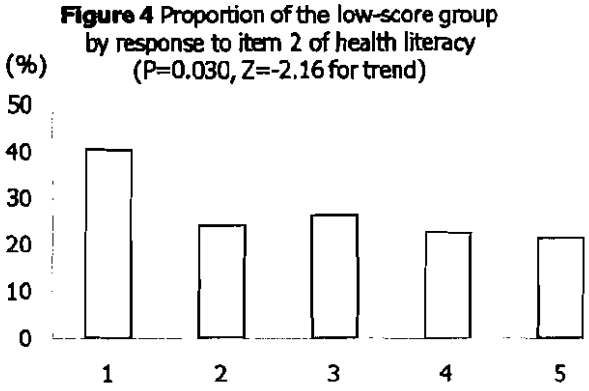


Figure 4. Proportion of participants with poor health knowledge, by response to item 2 ("Help reading") on the health literacy test. The question was, "How often do you have someone help you read hospital materials?" The 5-point Likert response scale was, "Always" (1), "Often" (2), "Sometimes" (3), "Occasionally" (4), or "Never" (5). Participants with lower literacy represented a higher proportion of those with a low score on the Japanese Health Knowledge Test.

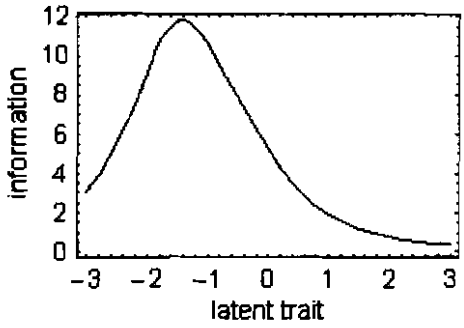


Figure 5. Item information function curve for the 15-item Japanese Health Knowledge Test.

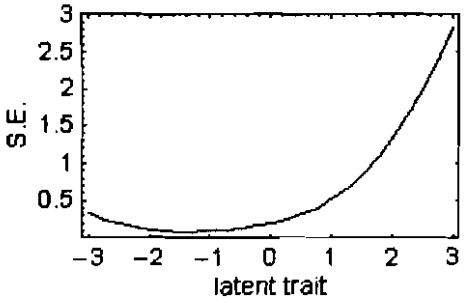


Figure 6. Standard error of the item information function for the 15-item Japanese Health Knowledge Test. S.E. indicates standard error.

Table 3. Score results of the 15-item Japanese Health Knowledge Test (n = 1040)

Characteristic	Total score			Group with low score[a]	
	Mean	SD	P-value	n (%)	P-value
Age (years)					
<65 (n = 685)	4.67	1.63	0.660[d]	162 (24)	0.996[b]
≥65 (n = 355)	4.63	1.57	(0.441)	84 (24)	(0.001)
Sex					
Male	4.62	1.63	0.455[d]	125 (25)	0.277[b]
Female	4.69	1.58	(0.748)	121 (22)	(1.181)
Income (Japanese Yen)					
<2 million	4.17	1.46	<0.001[d]	27 (29)	0.004[c]
2–3.99 million	4.55	1.60	(4.150)	72 (27)	(−2.89)
4–5.99 million	4.63	1.66		72 (25)	
6–7.99 million	4.73	1.54		32 (20)	
8 million or more	4.96	1.58		43 (18)	
Education					
<Grade 12	4.24	1.59	<0.001[d]	17 (33)	0.021[c]
High school graduate	4.50	1.55	(4.340)	98 (26)	(−2.31)
Vocational school	4.37	1.31		26 (24)	
Some college	4.78	1.66		31 (22)	
University or graduate degree	4.92	1.68		74 (20)	
Working status					
Working full-time	4.67	1.63	0.565[e]	110 (25)	0.938[b]
Homemaker	4.74	1.66	(0.740)	61 (22)	(0.805)
Working part-time	4.51	1.49		21 (23)	
Retired	4.71	1.55		30 (22)	
Currently not working	4.47	1.51		24 (25)	
Smoking					
Current	4.49	1.59	0.154[e]	56 (28)	0.079[b]
Former	4.79	1.53	(1.880)	47 (19)	(5.090)
Never	4.66	1.64		143 (24)	
Current alcohol use					
None/light	4.74	1.59	0.027[d]	121 (21)	0.004[c]
Moderate	4.57	1.61	(2.210)	109 (27)	(2.850)
Heavy	4.31	1.72		16 (36)	

SD = standard deviation.
[a]Participants with a score of 0–3 points, ie, mean − SD.
[b]The chi-square test was used. The numbers in parentheses are the chi-square statistic.
[c]The trend test was used. The numbers in parentheses are the z-statistic.
[d]Linear regression was used. The numbers in parentheses are the t-statistic.
[e]ANOVA was used. The numbers in parentheses are the F-statistic.

DISCUSSION

Using nominal categories modeling of item response theory analysis, we developed the 15-item J-HKT for Japanese adults. The instrument had a good ability to discriminate among those with poor health knowledge. In addition, items on the J-HKT and health literacy instruments were significantly correlated. The proportion of respondents with a low score on the J-HKT was higher among those with low literacy, which provides evidence of construct validity. Further, fully 36% of the participants had poor health knowledge (defined as a score of 0–3 of a possible 15 on the J-HKT). Finally, we found that poor health knowledge was

Table 4. Logistic regression analysis of the odds of a low score on the Japanese Health Knowledge Test (n = 1040)

Characteristic	Odds ratio	95% CI of odds ratio	P-value
Age	0.99	0.98–1.00	0.145
Sex			
Male[a]	1.00		
Female	0.91	0.66–1.25	0.557
Income (Japanese Yen)			
8 million or more[a]	1.00		
6–7.99 million	1.12	0.67–1.88	0.661
4–5.99 million	1.45	0.94–2.23	0.091
2–3.99 million	1.88	1.08–2.62	0.022
<2 million	1.84	1.02–3.31	0.042
Education			
University or graduate degree	1.00		
Soma college	1.26	0.76–2.08	0.366
Vocational school	1.18	0.69–2.02	0.538
High school graduate	1.43	0.99–2.06	0.058
<Grade 12	2.08	1.05–4.14	0.036
Current alcohol use			
None/light[a]	1.00		
Moderate	1.53	1.12–2.09	0.008
Heavy	2.28	1.16–4.47	0.017

[a]Reference group.
CI = confidence interval.

associated with low income, low educational attainment, and heavier current use of alcohol.

We used nominal categories modeling to elueidate the individual discriminating power and the effect of item position in the initial 48-item pool. This allowed us to identify items with good psychometric characteristics for inclusion in the 15-item J-HKT. Therefore, it is likely that we successfully developed a test that performs well in assessing health knowledge level among people with moderately poor health knowledge.

We chose to focus the discriminating capacity of this test at the low end of health knowledge, for several reasons. First, people with the lowest levels of health knowledge are those who have the worst health outcomes.[16,24,25] An increase in health knowledge among people who already have relatively greater knowledge is desirable, but is not likely to provide the biggest health impact. Next, focusing health resources on people with poor health knowledge is a means of minimizing health disparities.[26] People with poor health knowledge are likely to have more complex illnesses, and management of complex illnesses requires proper adherence to regimens via active patient involvement in treatment, which is more likely when illnesses are better understood.[5]

Several limitations should be noted. First, the results of our study were based on an online survey. A high proportion of Japanese adults use the internet, and while this mode of testing is much less expensive and much more convenient than in-person household interviews, it is possible that people in the sampling frame were younger, wealthier, and more educated than the general public.[27] As such, caution should be used in extrapolating our results.

Similarly, while the participation rate in this project is satisfactory for online research, it is likely that the participants were different from nonparticipants. Different methods for sampling the general population or patient populations with experience of frequent visits to clinicians (eg, due to chronic illness) might result in different distributions of J-HKT scores. There might also have been issues related to differential item functioning between participants and nonparticipants.[27] Although this paper presents a careful psychometric evaluation of the 15-item J-HKT, additional research is needed to ensure appropriate calibration.

Third, since this was an online survey, we do not know if the participants had help or discussed the questions with anyone else. The online panel registration system required a personal identification number and password, and did not allow participants to test more than once. However, participants had to read the questions, and poor reading skill may have resulted in an incorrect answer for an item that would have been answered correctly had it been read aloud. Further research in the form of a test-retest evaluation is needed to determine if the results of verbal administration differ from those of the written test.[28]

Fourth, based on the item information functions of the IRT analysis for each response to individual items, the content of several responses must be improved. For instance, on item 10, no participants selected responses 3 or 4, and, on item 1, nearly all participants selected response 3. Moreover, several items will require revision because of dynamic changes in the public's awareness of health information, due to rapid turnover in health-related knowledge in this era of rapid technological advance.

In summary, the current study described an online test of health knowledge among Japanese. We carefully evaluated the psychometric properties of this test and produced an instrument that can accurately discriminate among participants with poor health knowledge. The J-HKT is a convenient and valid measure of health knowledge, and can be used for the general Japanese public. Japanese public health practitioners and clinicians can easily use this quick test for the purposes of health education and disease prevention.

SUPPLEMENTS

How to answer the items below

Below are some medical terms that you may encounter or have encountered on various occasions in medical settings or situations. Please select the sentence that you think best describes each term. The objective of this test is to evaluate your awareness of healthcare terms. It is NOT a test to determine the number of correct responses. Please answer each Item based on your knowledge, even if you are unfamiliar with the terms.

Item 1 Please select the sentence that best describes the term "Tumor."

1. A state of cancer that can be life-threatening.
2. Early treatment, such as surgery, is necessary because it often metastasizes throughout the body.
3. A growth of tissue (mass of cells) that arises from abnormal cellular proliferation.
4. Growth is slow, and it does not spread to other parts of the body or invade surrounding tissue.

Item 2 Please select the sentence that best describes the term "Anti-tumor Drug."

1. It works for all forms of cancer, so it is given to almost all cancer patients.
2. Because this drug does not cure cancer, it is predominantly used for terminal cancer.
3. This drug suppresses cancer cell proliferation and eliminates cancer.
4. Due to its numerous adverse effects and limited therapeutic effect, this drug is used only when requested by patients.

Item 3 Please select the sentence that best describes the term "Ileus."

1. It has almost the same meaning as intestinal obstruction.
2. It does not occur to people who have had abdominal surgery in the past.
3. A condition where the passage of bowel contents is excessively rapid.
4. A small, sac-like protrusion that develops on the intestinal wall.

Item 4 Please select the sentence that best describes the term "Ulcer."

1. Because it is benign, there is no need to worry about cancer.
2. Duodenal ulcers may develop into cancer.
3. A condition where the surface of mucous membrane or skin is injured and deeply gouged.
4. Stomach ulcer usually heals on its own.

Item 5 Please select the sentence that best describes the term "Renal Failure."

1. Because it is asymptomatic and painless, treatment is generally not required.
2. A condition where the kidney is diseased and requires or almost requires dialysis (artificial kidney).
3. It is caused by chronic nephritis, not diabetes or hypertension.
4. It is caused by long-term, heavy alcohol consumption and causes jaundice (yellowish pigmentation of the skin).

Item 6 Please select the sentence that best describes the term "Influenza."

1. It is what we call the "common cold."
2. A bacterial infectious disease caused by the influenza bacteria.
3. It is 100% preventable by vaccine.
4. Antibiotics are ineffective.

Item 7 Please select the sentence that best describes the term "Arteriosclerosis."

1. Changes in the artery associated with old-age.
2. It is caused by diabetes and/or hypertension but progresses with age.
3. It is not caused by smoking.
4. It happens less in men than in women.

Item 8 Please select the sentence that best describes the term "Remission."

1. It is when an illness has been completely cured.
2. It is a phenomenon in which symptoms worsen due to chronic diseases.
3. It is when no further hospitalization or examination is necessary.
4. It is when symptoms are gone but the illness is not completely healed.

Item 9 Please select the sentence that best describes the term "Terminal Care."

1. Medical practice that emphasizes QOL enhancement more than life-sustaining treatment.
2. It is only for terminal cancer patients.
3. Medical services provided at train stations.
4. It refers to "Care of the Dying"

Item 10 Please select the sentenee that best describes the term "Hospice."

1. A hospital ward where once you enter, you never leave.
2. Hospitalization fees at a hospice cost more than fees at a regular hospital ward.
3. Palliative care is provided to ease physical, psychological and spiritual pain of terminally-ill patients.
4. A place for dying where no treatment is provided.

Item 11 Please select the sentence that best describes the term "Death with Dignity."

1. Administering a lethal injection for the purpose of stopping the heart and hastening death.
2. Choosing to die peacefully and naturally, maintaining one's dignity.
3. Committing suicide by ingesting poison.
4. It is when the patient refuses life-support for not wanting to cause his/her family any trouble.

Item 12 Please select the sentence that best describes the term "Clinical Pathway."

1. Comprehensive and standardized plan of care in which care categories, such as exam, surgery, administration of medication, treatment, nutrition, etc., are organized and sequenced over a specified course of time.
2. A schedule that specifies outpatient clinic physicians based on days and specialty.
3. An identification card required for hospital consultations.
4. Individually-developed care schedules that emphasize each physician's unique treatment protocol.

Item 13 Please select the sentence that best describes the term "Metabolic Syndrome."

1. An overweight person with greater-than-standard abdominal girth measurement.
2. An obese person who has a high level of "bad" cholesterol.
3. There is an increased risk of diabetes, hypertension, hyperlipidemia and complications due to accumulation of visceral fat.
4. Its cause is more due to heredity than life-style habits.

Item 14 Please select the sentence that best describes the term "EBM."

1. A standard medical practice that eliminates a physician's experience and instincts.
2. To practice medicine based on scientific evidence but also being considerate of each patient's situation and values.
3. To conduct research based on assumption and imagination.
4. To use treatment that has been reported to be effective in a small number of study cases.

Item 15 Please select the sentence that best describes the term "Evidence."

1. Treatment methods that are subjectively chosen and widely recommended by specialists.
2. A large majority of home remedies have "evidence" and is proven effective.
3. Treatment methods that have been proven effective in animal experiments.
4. Scientific evidence and proof that explain the effectiveness of treatment methods and medications.

ACKNOWLEDGEMENTS

We thank all the staff of the National Institute for Japanese Language and the Life Planning Center for their support of this research.

This study was supported by a Research Grant from Pfizer Health Research Funds. The funding source had no role in the collection or analysis of the study data.

REFERENCES

1. Nielsen-Bohlman L. Institute of Medicine. Health Literacy: A Prescription to End Confusion. National Academy Press; 2004.
2. Wolf MS, Gazmararian JA, Baker DW. Health literacy and functional health status among older adults. Arch Intern Med. 2005;165:1946–52.
3. Paasche-Orlow MK, Wolf MS. The causal pathways linking health literacy to health outcomes. Am J Health Behav. 2007;31 Suppl 1:S19–26.
4. Gazmararian JA, Williams MV, Peel J, Baker DW. Health literacy and knowledge of chronic disease. Patient Educ Couns. 2003;51:267–75.
5. Wolf MS, Davis TC, Cross JT, Marin E, Green K, Bennett CL. Health literacy and patient knowledge in a Southern US HIV clinic. Int J STD AIDS. 2004;15:747–52.
6. Fang MC, Machtinger EL, Wang F, Schillinger D. Health literacy and anticoagulation-related outcomes among patients taking warfarin. J Gen Intern Med. 2006;21:841–6.
7. Kalichman SC, Ramachandran B, Catz S. Adherence to combination antiretroviral therapies in HIV patients of low health literacy. J Gen Intern Med. 1999;14:267–73.
8. Schillinger D, Grumbach K, Piette J, et al. Association of health literacy with diabetes outcomes. JAMA. 2002;288:475–82.
9. Kripalani S, Henderson LE, Chiu EY, Robertson R, Kolm P, Jacobson TA. Predictors of medication self-management skill in a low-literacy population. J Gen Intern Med. 2006;21:852–6.
10. Gazmararian JA, Kripalani S, Miller MJ, Echt KV, Ren J, Rask K. Factors associated with medication refill adherence in cardiovascular-related diseases: a focus on health literacy. J Gen Intern Med. 2006;21:1215–21.
11. 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;40:395–404.
12. Sudore RL, Mehta KM, Simonsick EM, et al. Limited literacy in older people and disparities in health and healthcare access. J Am Geriatr Soc. 2006;54:770–6.
13. Howard DH, Sentell T, Gazmararian JA. Impact of health literacy on socioeconomic and racial differences in health in an elderly population. J Gen Intern Med. 2006;21:857–61.
14. 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;172:980–6.
15. Rothman R, Malone R, Bryant B, Horlen C, DeWalt D, Pignone M. The relationship between literacy and glycemic control in a diabetes disease-management program. Diabetes Educ. 2004;30:263–73.
16. Rothman RL, Malone R, Bryant B, et al. The Spoken Knowledge in Low Literacy in Diabetes scale: a diabetes knowledge scale for vulnerable patients. Diabetes Educ. 2005;31:215–24.
17. 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;23:561–6.
18. Wallace LS, Rogers ES, Roskos SE, Holiday DB, Weiss BD. Brief report: screening items to identify patients with limited health literacy skills. J Gen Intern Med. 2006;21:874–7.
19. Chew LD, Bradley KA, Boyko EJ. Brief questions to identify patients with inadequate health literacy. Fam Med. 2004;36:588–94.
20. Bock R. Estimating item parameters and latent ability when responses are scored in two or more nominal categories. Psychometrika. 1972;39:29–51.
21. Thissen D, Steinberg L. A Response Model for Multiple-Choice Items. In: Linden W, Hambleton R, eds. Handbook of Modern Item Response Theory. New York, NY: Springer; 1997.
22. Okubo T. An item parameter estimation programme for nominal categories model using R. DNC research note. 2007;RN-07-18.
23. Cuzick JA. Wilcoxon-type test for trend. Stat Med. 1985;4:87–90.
24. Kleinbeck C. Reaching positive diabetes outcomes for patients with low literacy. Home Healthc Nurse. 2005;23:16–22.
25. Olayemi SO, Oreagba IA, Akinyede A, Adepoju GE. Educational intervention and the health seeking attitude and adherence to therapy by tuberculosis patients from an urban slum in lagos Nigeria. Niger Postgrad Med J. 2009;16:231–5.
26. Marmot M. Social determinants of health inequalities. Lancet. 2005;365:1099–104.
27. Stewart AL, Napoles-Springer AM. Advancing health disparities research: can we afford to ignore measurement issues? Med Care. 2003;41:1207–20.
28. Tokuda Y, Doba N, Butler JP, Paasche-Orlow MK. Health literacy and physical and psychological wellbeing in Japanese adults. Patient Educ Couns. 2009;75:411–7.

Journal of Substance Abuse Treatment 39 (2010) 378–383

Journal of Substance Abuse Treatment

Brief article

Physician introduction to opioids for pain among patients with opioid dependence and depressive symptoms

Judith I. Tsui,[a,*] Debra S. Herman,[b,c] Malyna Kettavong,[b] Daniel Alford,[a] Bradley J. Anderson,[b] Michael D. Stein[b]

[a]*Boston University School of Medicine/Boston Medical Center, Boston MA 02118*
[b]*Butler Hospital, 345 Blackstone Blvd, Providence, RI 02906*
[c]*Warren Alpert School of Medicine at Brown University, Providence, RI, 02912*

Received 16 April 2010; received in revised form 25 June 2010; accepted 25 June 2010

Abstract

This study determined the frequency of reporting being introduced to opioids by a physician among opioid-dependent patients. Cross-sectional analyses were performed using baseline data from a cohort of opioid addicts seeking treatment with buprenorphine. The primary outcome was a response to the question: "Who introduced you to opiates?" Covariates included sociodemographics, depression, pain, and current and prior substance use. Of 140 participants, 29% reported that they had been introduced to opioids by a physician. Of those who were introduced to opioids by a physician, all indicated that they had initially used opioids for pain, versus only 11% of those who did not report being introduced to opioids by a physician ($p < .01$). There was no difference in current pain (78% vs. 85%, $p = .29$); however, participants who were introduced to opioids by a physician were more likely to have chronic pain (63% vs. 43%, $p = .04$). A substantial proportion of individuals with opioid dependence seeking treatment may have been introduced to opioids by a physician.

Keywords: Opioid dependence; Pain; Physician

1. Introduction

Based on 2008 national data, nearly 2 million individuals in the Unite States abuse or depend on opioids, with most reporting abuse of prescription opioids (Substance Abuse and Mental Health Services Administration, 2009). Individuals may be introduced to opioids through a variety of nonmedical and medical situations. Patients who are introduced to opioids through a physician may subsequently develop abuse or dependence either through continued use of prescription opioids or heroin. However, most persons who have misused prescription opioids report receiving them from a friend or relative (Substance Abuse and Mental Health Services Administration, 2009), and others may purchase diverted opioids from other sources (Inciardi, Surratt, Cicero, & Beard, 2009). Therefore, prescription opioid abuse and dependence may develop without provision of script from a physician. At the same time, physician prescribing rates for opioids are increasing (Caudill-Slosberg, Schwartz, & Woloshin, 2004; Gilson, Ryan, Joranson, & Dahl, 2004; Olsen, Daumit, & Ford, 2006; Sullivan et al., 2008), raising concern that the medical providers may share the responsibility for the rise in opioid abuse and dependence.

The risk of developing opioid abuse or dependence after being prescribed opioids for acute or chronic pain by a physician is unknown. A meta-analysis of iatrogenic addiction (defined as an "addiction of a patient to a drug initially prescribed for a medical condition") concluded that the literature could not establish whether the risk was high (>10%) or low (<0.1%) (Wasan, Correll, Kissin, O'Shea, & Jamison, 2006). However, the practice of prescribing opioids for pain is fairly common: research suggest that up to one

* Corresponding author. 801 Massachusetts Ave, Rm 2078, Boston, MA 02118. Tel.: +1 617 414 6912; fax: +1 617 414 4676.
E-mail address: judith.tsui@bmc.org (J.I. Tsui).

0740-5472/10/$ – see front matter
doi:10.1016/j.jsat.2010.06.012

fifth of adults are prescribed opioids annually (MMWR, 2010; Williams, Sampson, Kalilani, Wurzelmann, & Janning, 2008). Even if the true risk is low, the widespread practice of prescribing opioids could still result in a substantial absolute number of cases of opioid abuse and dependence, thereby contributing to the problem. Therefore, the relatively common practice of prescribing opioids could still result in a substantial absolute number of cases of opioid abuse and dependence. Furthermore, it is important to examine the phenomenon of physician introduction to opioids among a population of prescription and nonprescription opioid users because (a) patients who were initially introduced to opioids by a physician may have shifted their use to heroin and (b) prescription opioids can be procured without use of a physician script.

This study examined the prevalence of reporting being introduced to opioids by a physician among a cohort of patients with opioid dependence who were seeking addiction treatment with buprenorphine. In addition, we performed descriptive analyses to understand how participants who reported opioid introduction through a physician differed from those who did not. We hypothesized that opioid-dependent participants who had been introduced by a physician would be more likely to use prescription opioids as opposed to heroin and would be more likely to have pain because this would be an underlying reason for an initial introduction to opioids by a physician.

2. Methods

2.1. Study sample and design

This cross-sectional study used baseline data from participants in a randomized, controlled trial to determine whether treatment for depressive symptoms increases treatment retention among opioid-dependent patients initiating buprenorphine (Stein et al., 2010). Participants were recruited through community advertising, physician referrals, and word-of-mouth. Study inclusion criteria included the following: age 18–65, *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition* diagnosis of opioid dependence, score on the Modified Hamilton Depression Revised Scale greater than 14 (Miller, Bishop, Norman & Maddover, 1985), 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, or paranoid disorder), no currently prescribed medications for depression (participants were not specifically excluded if they were taking a tricyclic antidepressant only 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, 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 the parent study. Seven participants (4.8%) were missing data on the question that asked whether a physician had introduced them to opiates and were excluded from the analyses, leaving a final study cohort of 140.

2.2. Measures

The primary outcome examined was participants' responses to the question: "Who introduced you to opiates?" Possible responses included physician, sexual partner, friend, family member, stranger, and no one. Covariates included the demographic variables age, gender, race (White vs. non-White), current (past 30 day) employment, and insurance status. Clinical variables included severe depression, current pain, chronic pain, self-report of starting opioids for pain, having a primary care provider, and regular use of alcohol, marijuana, or cocaine prior to opioid use. Severe depression was defined as a score greater than 28 on the Beck Depression Inventory II (Beck, Steer, & Brown, 1996). Current pain was defined as having any pain in the past week; chronic pain was defined as pain that had been present for at least 6 months. Current pain was rated in severity by using the Visual Analog Scale (VAS; Melzack, 1987). Chronic severe pain was defined as pain experienced 6 months or longer that (a) caused at least moderate interference (defined by a single question from the Short Form-12 Questionnaire (Ware, Kosinski, & Keller, 1996) or (b) was of at least moderate in intensity (defined as a score above the median VAS score for the group). This definition was adapted from a study of chronic severe pain in methadone maintenance patients by the Rosenblum et al. (2003). 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?" Information on current (last 30 days) and past use of prescription opioids and heroin (including route of administration) was obtained using the Addiction Severity Index (McLellan et al., 1992). Regular use of alcohol, marijuana, and cocaine was determined by the question: "Prior to starting opiates, did you ever have daily or regular use of (drug)?"

2.3. Statistical analysis

All analyses were performed using baseline study data. The prevalence of physician introduction to opioids was determined by calculating the proportion with that report from the total sample. Descriptive analyses were performed comparing individuals who reported a physician introduction to opioids to those who did not report a physician introduction. We examined differences in demographic, clinical, and substance use-related variables between participants using Student's *t* tests and Pearson's chi-square tests.

Table 1
Characteristics of opioid-dependent patients who do and do not report a physician introducing them to opioids

Characteristics	Physician Introduced (n = 40) % or M (±SD)	Physician Did Not Introduce (n = 100) % or M (±SD)	p
Age	38 (±9)	37 (±10)	.87
Female	28	22	.49
Non-White	15	22	.35
Married	10	7	.52
Employed	45	56	.24
Has health insurance	44	33	.24
Has a primary care provider	38	37	.96
Duration of opioid use	9 (±8)	9 (±7)	.74
Prior injection drug use	38	76	<.01
Regular use of alcohol prior to opioids	35	34	.91
Regular use of marijuana prior to opioids	53	72	.03
Regular use of cocaine prior to opioids	23	45	.01

All statistical analyses were conducted using Stata version 10.0 (College Station, TX).

3. Results

Of the 140 opioid-dependent participants seeking treatment in the sample, 40 (29%) reported that they had been introduced to opioids by a physician. The mean age in the sample was 38 years (SD ± 10), 24% were female, and the average duration of opioid use was 9 years. There were no significant differences in gender, age, race, marital status, employment, or insurance status among individuals who did and did not report being introduced to opioids by a physician (Table 1). Individuals who were introduced to opioids through a physician were less likely to have injected drugs. Regular use of alcohol prior to starting opioids was equally reported among those who were and were not introduced by a physician to opioids. However, individuals who were introduced by a physician were significantly less likely to report prior use of marijuana and cocaine.

All of the participants who reported being introduced to opioids by a physician indicated that they had started using opioids to treat pain as compared to only 11% of those not introduced to opioids by a physician (Table 2). The prevalence of any current or chronic pain was high in the overall sample (83% and 49%, respectively). There were no significant differences in the proportions reporting current pain (any pain in the past week) between individuals who had and had not been initially introduced to opioids by a physician (Table 2). Among individuals who reported pain, the mean VAS score between the two groups did not significantly differ (57 [SD ± 4] for physician introduced vs. 60 [SD ± 3] for not physician introduced; p = .57). Individuals who had been introduced to opioids by a physician were more likely to have chronic pain and chronic severe pain. The percentage of individuals with severe depression did not differ between the two groups. Participants who were introduced to opioids by a physician were more likely to admit to using a physician to procure a prescription for opioids to get high with and to seek out a physician who "gives opiate prescriptions without asking too many questions."

Participants introduced to opioids by a physician were more likely to be currently using prescription opioids only (Fig. 1) compared with participants who were not introduced by a physician (chi-square, p < .01). However, 32% of opioid-dependent individuals who were physician-introduced reported currently using heroin in combination with prescription opioids, and 10% were using heroin exclusively. Likewise, participants who were introduced to opioids by a

Table 2
Pain, depression, and drug-related aberrant behaviors among opioid-dependent patients who do and do not report a physician introducing them to opiates

Behaviors	Physician Introduced (n = 40), %	Physician did not introduce (n = 100), %	p
Introduced to opioids for pain	100	11	<.01
Current (past week) pain	78	85	.29
Chronic (>6 months) pain	63	43	.04
Chronic severe pain	53	30	.01
Severe Depression	38	48	.26
Regular opioid use was prescribed by a physician	85	14	<.01
Ever used a physician to procure a prescription for opioids to get high	50	32	.05
Ever used a physician to procure a prescription for opioids to sell to others	10	4	.17
Ever sought out a physician known to easily prescribe opioids	33	17	.04

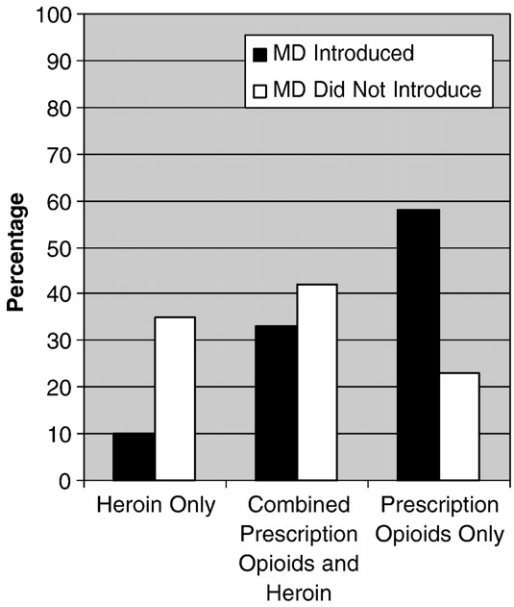


Fig. 1. Patterns of current opioid use among opioid-dependent participants who were and were not introduced to opioids by a physician.

physician were half as likely to currently inject drugs (28% vs. 57%, $p < .01$).

4. Discussion

Opioids are prescribed in medical settings for acute and chronic pain, and physicians may provide an introduction to opioids in individuals who subsequently develop opioid abuse and dependence. This study found that among a cohort of opioid-dependent patients who were seeking addiction treatment with buprenorphine, 29% reported that they had been introduced to opioids through a physician treating their pain. To our knowledge, this is the first study to assess this question in a population of opioid-dependent patients who were users of both prescription and nonprescription opioids. This study, although it does not define the risk of developing opioid abuse or dependence among patients who are prescribed opioids in a medical setting, provides insights on the fraction of patients whose opioid dependence can be linked to a medical introduction to opioids.

This study expands the current understanding of physicians' roles in the growing trend of opioid abuse and dependence. Prescription opioid misuse is reportedly increasing (Gilson et al., 2004; Zacny et al., 2003), as are overdose deaths related to prescription opioids (Hall et al., 2008; Warner, Chen, & Makuc, 2009). Concurrently, physician prescribing rates for opioids are increasing (Caudill-Slosberg et al., 2004; Gilson et al., 2004; Olsen et al., 2006; Sullivan et al., 2008), raising concern that the medical providers may share responsibility for the current trend. Physicians are encouraged to screen for and treat pain, and opioids are effective in treating pain (Ballantyne & Mao, 2003), although the risk for subsequent addiction is not known (Wasan et al., 2006). Even if the risk is low, the widespread practice of prescribing opioids among physicians could still result in a substantial absolute number of cases of opioid dependence and thereby contribute to the problem of opioid dependence. This study suggests that physician introduction to opioids is a factor for a sizeable proportion of individuals who have opioid dependence requiring addiction treatment.

The finding that participants who reported being introduced to opioids through a physician were more likely to have chronic pain is consistent with prior research. Studies of opioid-dependent patients on methadone maintenance show that more than a third report chronic pain (Rosenblum et al., 2003; Rosenblum et al., 2007). At least two prior studies have found a higher prevalence of chronic pain among opioid-dependent patients who were primarily prescription drug users compared to nonprescription drug users (Brands, Blake, Sproule, Gourlay, & Busto, 2004; Rosenblum et al., 2007). However, it is interesting to note that the proportion of individuals with current pain was similar among those who did and did not report a physician introduction to opioids. It is possible that some participants who did not initiate opioids for pain may have developed pain for reasons related to their opioid use. Such causes for pain in this cohort might include injuries from trauma, depression (Bair, Robinson, Katon, & Kroenke, 2003), chronic medical conditions that are associated with pain such as hepatitis C virus (Barkhuizen et al., 1999; Silberbogen, Janke, & Hebenstreit, 2007; Thompson & Barkhuizen, 2003), or opioid-induced hyperalgesia (Chu, Angst, & Clark, 2008).

Our study cohort included both current users and nonusers of prescription opioids, and results demonstrated that most individuals who were introduced to opioids by a physician for pain continued to use prescription opioids over heroin. Still, 42% of participants who reported that a physician introduced them to opioids were currently using some heroin (either exclusively or in combination with prescription opioids), and more than a quarter were injecting opioids, which confirms prior research that suggest shifting trajectories of opioid abuse among individuals who are initially prescribed opioids. A study of opioid-dependent patients in a methadone treatment program in Canada found that 24% of participants reported using prescription opioids (medically or nonmedically) initially and heroin later (Brands et al., 2004), and a study of 72 methadone maintenance treatment programs in the United States found that 69% of primary heroin users had ever used prescription opioids (Rosenblum et al., 2007). Our finding that opioid-dependent individuals who are initially introduced to opioids for pain are more likely to admit to prescription drug aberrant behaviors (telling a doctor they had pain to obtain opioids so that they could get high, or seeking out a physician known to prescribe opioids "without asking too many questions")

suggests that the patients who display drug aberrant behaviors are more likely to have developed their opioid addiction in the context of pain management through a physician, as opposed to initiating use through illicit opioids. This may be relevant to dispelling the bias against use of effective pain medicines for injection drug users (Breitbart et al., 1996).

There are several limitations to this study. The relatively small sample size limited power to detect statistically significant differences. The study questionnaire asked whether the participant was introduced to opioids by a physician but did not obtain detailed information on the setting (e.g., office or emergency department), the respondent's ongoing relationship with that prescribing physician, or the amount and duration of the initial opioid prescription. Furthermore, this question relied on patient recall, which could be subject to bias in either direction, and patients' conceptions of what constitutes an "introduction" to opiates may have differed. Second, our study was conducted among a sample of opioid-dependent individuals with depressive symptoms who were seeking treatment with buprenorphine, which may limit generalizability. However, depression is common among opioid-addicted patients: Studies estimate that approximately a third to one half have depression (Brienza et al., 2000; Croughan, Miller, Wagelin, & Whitman, 1982; Khantzian & Treece, 1985; Rounsaville, Weissman, Crits-Christoph, Wilber, & Kleber, 1982). Furthermore, our sample appears to be similar to other populations of buprenorphine-treated patients with regard to the proportion currently using heroin versus nonheroin opioids, supporting its overall general representativeness (Stanton, McLeod, Luckey, Kissin, & Sonnenfeld). Finally, by excluding individuals on antidepressants, which may include a small subset on tricyclic antidepressants for depression and pain, ours may be an underestimate of the true proportion of opioid-dependent individuals who are introduced to opioids by a physician for pain.

In summary, this study found that 29% of patients who were presenting for treatment for opioid dependence reported that a physician had introduced them to opioids. This finding reinforces the need for physicians to carefully select patients before initiating an opioid therapeutic trial and, once prescribing, to monitor patients regularly for opioid misuse, abuse, and dependence (Passik, 2009). More prospective research is needed to examine the trajectories of opioid use that occur among individuals who introduced to opioids through a physician and to develop strategies to prevent patients from developing addiction to opioids.

Acknowledgments

This study was funded by the National Institute on Drug Abuse (NIDA; DA 022207), Clinical Trial # NCT 00475878. Dr. Stein is a recipient of a NIDA Mid-Career Award (DA 000512).

References

MMWR. (2010). Adult use of prescription opioid pain medications—Utah, 2008. *MMWR Morbidity and Mortal Weekly Report*, *59*, 153–157.

Bair, M. J., Robinson, R. L., Katon, W., & Kroenke, K. (2003). Depression and pain comorbidity: A literature review. *Archives of Internal Medicine*, *163*, 2433–2445.

Ballantyne, J. C., & Mao, J. (2003). Opioid therapy for chronic pain. *New England Journal of Medicine*, *349*, 1943–1953.

Barkhuizen, A., Rosen, H. R., Wolf, S., Flora, K., Benner, K., & Bennett, R. M. (1999). Musculoskeletal pain and fatigue are associated with chronic hepatitis C: A report of 239 hepatology clinic patients. *American Journal of Gastroenterology*, *94*, 1355–1360.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the BDI-II.* San Antonio, TX: The Psychological Corporation.

Brands, B., Blake, J., Sproule, B., Gourlay, D., & Busto, U. (2004). Prescription opioid abuse in patients presenting for methadone maintenance treatment. *Drug and Alcohol Dependence*, *73*, 199–207.

Breitbart, W., Rosenfeld, B. D., Passik, S. D., McDonald, M. V., Thaler, H., & Portenoy, R. K. (1996). The undertreatment of pain in ambulatory AIDS patients. *Pain*, *65*, 243–249.

Brienza, R. S., Stein, M. D., Chen, M., Gogineni, A., Sobota, M., Maksad, J., et al. (2000). Depression among needle exchange program and methadone maintenance clients. *Journal of Substance Abuse Treatment*, *18*, 331–337.

Caudill-Slosberg, M. A., Schwartz, L. M., & Woloshin, S. (2004). Office visits and analgesic prescriptions for musculoskeletal pain in US: 1980 vs. 2000. *Pain*, *109*, 514–519.

Chu, L. F., Angst, M. S., & Clark, D. (2008). Opioid-induced hyperalgesia in humans: Molecular mechanisms and clinical considerations. *Clinical Journal of Pain*, *24*, 479–496.

Croughan, J. L., Miller, J. P., Wagelin, D., & Whitman, B. Y. (1982). Psychiatric illness in male and female narcotic addicts. *Journal of Clinical Psychiatry*, *43*, 225–228.

Gilson, A. M., Ryan, K. M., Joranson, D. E., & Dahl, J. L. (2004). A reassessment of trends in the medical use and abuse of opioid analgesics and implications for diversion control: 1997–2002. *Journal of Pain Symptom Management*, *28*, 176–188.

Hall, A. J., Logan, J. E., Toblin, R. L., Kaplan, J. A., Kraner, J. C., Bixler, D., et al. (2008). Patterns of abuse among unintentional pharmaceutical overdose fatalities. *JAMA*, *300*, 2613–2620.

Inciardi, J. A., Surratt, H. L., Cicero, T. J., & Beard, R. A. (2009). Prescription opioid abuse and diversion in an urban community: The results of an ultrarapid assessment. *Pain Medicine*, *10*, 537–548.

Khantzian, E. J., & Treece, C. (1985). DSM-III psychiatric diagnosis of narcotic addicts. Recent findings. *Archive of General Psychiatry*, *42*, 1067–1071.

McLellan, A. T., Kushner, H., Metzger, D., Peters, R., Smith, I., Grissom, G., et al. (1992). The Fifth Edition of the Addiction Severity Index. *Journal of Substance Abuse Treatment*, *9*, 199–213.

Melzack, R. (1987). The short-form McGill Pain Questionnaire. *Pain*, *30*, 191–197.

Miller, I., Bishop, S., Norman, W., & Maddever, H. (1985). The Modified Hamilton Rating Scale for Depression: Reliability and validity. *Psychiatry Research*, *14*, 131–142.

Olsen, Y., Daumit, G. L., & Ford, D. E. (2006). Opioid prescriptions by U.S. primary care physicians from 1992 to 2001. *J Pain*, *7*, 225–235.

Passik, S. D. (2009). Issues in long-term opioid therapy: Unmet needs, risks, and solutions. *Mayo Clin Proc*, *84*, 593–601.

Rosenblum, A., Joseph, H., Fong, C., Kipnis, S., Cleland, C., & Portenoy, R. K. (2003). Prevalence and characteristics of chronic pain among chemically dependent patients in methadone maintenance and residential treatment facilities. *JAMA*, *289*, 2370–2378.

Rosenblum, A., Parrino, M., Schnoll, S. H., Fong, C., Maxwell, C., Cleland, C. M., et al. (2007). Prescription opioid abuse among enrollees into methadone maintenance treatment. *Drug and Alcohol Dependence*, *90*, 64–71.

Rounsaville, B. J., Weissman, M. M., Crits-Christoph, K., Wilber, C., & Kleber, H. (1982). Diagnosis and symptoms of depression in opiate addicts. Course and relationship to treatment outcome. *Archive of General Psychiatry*, *39*, 151−156.

Silberbogen, A. K., Janke, E. A., & Hebenstreit, C. (2007). A closer look at pain and hepatitis C: Preliminary data from a veteran population. *Journal of Rehabilitation Research and Development*, *44*, 231−244.

Stanton, L., McLeod, C., Luckey, B., Kissin, W. B., & Sonnenfeld, L. J. SAMHSA/CSAT Evaluation of the Buprenorphine Waiver Program. http://buprenorphine.samhsa.gov/ASAM_06_Final_Results.pdf. accessed 1/22/09.

Stein, M. D., Herman, D. S., Kettavong, M., Cioe, P. A., Friedmann, P. D., Tellioglu, T., et al. (2010). Antidepressant treatment does not improve buprenorphine retention among opioid-dependent persons. *Journal of Substance Abuse Treatment*, *39*, 157−166.

Substance Abuse and Mental Health Services Administration. (2009). Results from the 2008 National Survey on Drug Use and Health: National Findings. Office of Applied Studies, NSDUH Series H-36, HHS Publication No. SMA 09-4434, Rockville, MD.

Sullivan, M. D., Edlund, M. J., Fan, M. Y., Devries, A., Brennan Braden, J., & Martin, B. C. (2008). Trends in use of opioids for non-cancer pain conditions 2000–2005 in commercial and Medicaid insurance plans: The TROUP study. *Pain*, *138*, 440−449.

Thompson, M. E., & Barkhuizen, A. (2003). Fibromyalgia, hepatitis C infection, and the cytokine connection. *Current Pain and Headache Report*, *7*, 342−347.

Ware, J. E., 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.

Warner, M., Chen, L. H., & Makuc, D. M. (2009). Increase in fatal poisonings involving opioid analgesics in the United States, 1999–2006. NCHS data brief(No. 22), Hyattsville, MD: National Center for Health Statistics. 2009.

Wasan, A. D., Correll, D. J., Kissin, I., O'Shea, S., & Jamison, R. N. (2006). Iatrogenic addiction in patients treated for acute or subacute pain: A systematic review. *Journal of Opioid Management*, *2*, 16−22.

Williams, R. E., Sampson, T. J., Kalilani, L., Wurzelmann, J. I., & Janning, S. W. (2008). Epidemiology of opioid pharmacy claims in the United States. *Journal of Opioid Management*, *4*, 145−152.

Zacny, J., Bigelow, G., Compton, P., Foley, K., Iguchi, M., & Sannerud, C. (2003). College on Problems of Drug Dependence taskforce on prescription opioid non-medical use and abuse: Position statement. *Drug and Alcohol Dependence*, *69*, 215−232.

Improving Decision Making at the End of Life With Video Images

Angelo E. Volandes, MD, Michael J. Barry, MD, Yuchiao Chang, PhD, Michael K. Paasche-Orlow, MD, MA, MPH

Background. *Decision making at the end of life is frequently complex and often filled with uncertainty. We hypothesized that people with limited health literacy would have more uncertainty about end-of-life decision making than people with adequate literacy. We also hypothesized that video images would decrease uncertainty.* ***Design.*** *Before and after oral survey.* ***Participants.*** *Subjects presenting to their primary care physicians.* ***Methods.*** *Subjects were asked about their preferences for end-of-life care after they heard a verbal description of advanced dementia and were asked to rate the level of their uncertainty. Subjects then viewed a video of a patient with advanced dementia and were asked again about their preferences and uncertainty. Uncertainty was measured using the Decisional Conflict Scale with score ranges from 3 (***high uncertainty***) to 15 (***no uncertainty***). Health literacy was measured using the Rapid Estimate of Adult Literacy in Medicine, and subjects were divided into 3 literacy categories: low (0–45, 6th grade and below), marginal (46–60, 7th–8th grade), and adequate (61–66, 9th grade and above).* ***Results.*** *A total of 146 patients completed the interview. Prior to the video, the average uncertainty scores for subjects with low, marginal, and adequate health literacy were 10.8, 12.4, and 13.5, respectively (*$P < 0.0001$*). After the video, the 3 groups had similar uncertainty about their decisions. The average uncertainty scores for subjects with low, marginal, and adequate health literacy were 13.6, 14.1, and 14.5, respectively (*$P = 0.046$*).* ***Conclusions.*** *Subjects with limited health literacy expressed more uncertainty about their preferences for end-of-life care than did subjects with adequate literacy. Our video decision aid improved end-of-life decision making by decreasing uncertainty regarding subjects' preferences, especially for those with limited literacy.* ***Key words:*** *end-of-life decision making; video; decision aids; uncertainty; advance care planning; health literacy; ethics; palliative care; dementia; Alzheimer's disease.* ***(Med Decis Making 2010;30:29–34)***

Decision making at the end of life is frequently complex and often filled with uncertainty.[1] Patients are increasingly asked to describe their preferences for medical care in advance of having a disease such as dementia.[2,3] Physicians use verbal or written descriptions to communicate future health states, but serious doubts have been raised regarding whether patients can realistically envision future health states on the basis of verbal descriptions alone.[4]

Received 16 October 2007 from the General Medicine Unit, Massachusetts General Hospital, Boston (AEV, MJB, YC); Harvard Medical School, Boston, Massachusetts (AEV, MJB, YC); and Department of Medicine, Boston University School of Medicine, Boston, Massachusetts (MKP-O). Revision accepted for publication 28 February 2009.

Address correspondence to Angelo E. Volandes, MD, Massachusetts General Hospital, General Medicine Unit, 50 Staniford Street, 9th floor, Boston, MA 02114; e-mail: avolandes@partners.org.

DOI: 10.1177/0272989X09341587

Limited health literacy is a prevalent barrier to accurate communication that can pose significant obstacles to informed decision making at the end of life.[5] Although more than 90 million American adults have limited health literacy,[6,7] techniques that can be used to overcome communication barriers posed by limited health literacy are frequently not employed. For example, visual images have been shown to improve communication of information with patients.[8–11] The medium of video allows patients to envision health states in a manner not easily captured with verbal communication and can both engage people and efficiently communicate information about the experience of illness.[12]

One measure of improved decision making at the end of life includes decreasing uncertainty regarding decisions.[1] We hypothesized that people with limited health literacy would have more uncertainty about end-of-life decision making than people with adequate literacy. In addition, we hypothesized that

video images would decrease uncertainty and that this improvement would be greatest for people with limited health literacy. We conducted a before-and-after trial to examine whether a video of a patient with advanced dementia could improve decision making by decreasing uncertainty.

METHODS

Participants

Patients at 6 study sites who were older than 40 y and were scheduled to see a general internist were eligible to participate. The study sites consisted of urban and suburban primary care clinics affiliated with 2 teaching hospitals in the greater Boston area. Subjects were excluded if they had previously had a close relationship with a person with advanced dementia because they would likely have had 1st-hand experience and knowledge of the disease. Subjects were also excluded if, in the judgment of the physician interviewer, they lacked medical decision-making capacity at the time of the interview or if they did not speak English, the language of our validated tools. Each interview was conducted by 1 of 2 physicians between 1 December 2005 and 31 January 2007. Approval for the project was granted by the Institutional Review Boards of the affiliated hospitals.

Design

A structured questionnaire was developed for this study. Survey questions were generated from a review of the medical ethics literature and consultations with medical ethics, palliative care, geriatric, and neurology experts. Early versions of the survey were pretested with subjects recruited from primary care clinics.

After obtaining verbal informed consent, interviewers defined advanced dementia in simple language, highlighting functional impairments based on the Functional Assessment Staging[13] criteria, including inability to communicate understandably with others, inability to ambulate without assistance, and inability to feed oneself (see the Appendix for text of the verbal description). We then outlined 3 levels of medical treatments and the goals associated with each level: life-prolonging care, limited care, and comfort care.[12] Subjects were then asked 3 questions measuring uncertainty regarding their preferences.

Each subject next viewed a 2-min video of a patient with advanced dementia. The video depicts the salient features of advanced dementia. The narrative that accompanies the video is identical to the verbal description used to assess subjects' initial preferences. The design, content, and structure of the video intervention were reviewed for accuracy by 3 geriatricians and 5 neurologists, all of whom specialize in the care of patients with dementia. The video was also designed with the close collaboration of caregivers of patients with advanced dementia. (The film clip is available online at www.ACPdecisions.com.) Each subject was then asked the same questions as before regarding preferences and uncertainty.

Measures

Answers to sociodemographic questions (age, race, gender, etc.) were self-reported by the subjects. Health literacy was measured at the end of the interview using the validated Rapid Estimate of Adult Literacy in Medicine tool (REALM).[14] The survey is available on request.

The main outcome measures were the difference of uncertainty regarding individual preferences for medical care across the 3 levels of health literacy before and after viewing a video of advanced dementia. Our measure for uncertainty was the uncertainty subscale of the Decisional Conflict Scale.[15] This validated subscale uses 3 questions to elicit the level of uncertainty regarding a subject's decision making. The subscale ranges from a score of 3 (*high uncertainty*) to a score of 15 (*no uncertainty*).

Our measure for health literacy was the 66-word REALM.[14] This is a 2- to 3-min English test of medically relevant vocabulary. The REALM is a valid test of word pronunciation and has been shown to correlate well with tests that evaluate a range of literacy skills. As others have done, we defined 3 categories for literacy: low literacy (REALM score of 0–45, 6th grade and below), marginal literacy (REALM score of 45–60, 7th–8th grade), and adequate literacy (REALM score of 61–66, 9th grade and above).[7]

Statistical Analysis

Means, standard deviations, and 95% confidence intervals were used to summarize the uncertainty scores before and after the video as well as the changes in uncertainty. Changes in preferences from before to after the video were analyzed with McNemar's test. Analysis of variance techniques were used to compare the prevideo uncertainty level, postvideo uncertainty level, and the change in uncertainty level among 3 ordered categories of health literacy in 3 separate models.

Multivariable regression analyses were used to examine the effect of health literacy on level of uncertainty, adjusting for other potential confounding factors. Factors significant at univariate $P < 0.20$ were tested in these models. Our independent variable of interest, health literacy, was designated to remain in each of our models. We further excluded variables that were highly associated with the variable of interest, health literacy, using χ^2 tests for association. Two-sided P values <0.05 were considered statistically significant. Data were analyzed using SAS software, version 9.1 (SAS Institute, Cary, NC).

RESULTS

Study Participants

A total of 146 subjects were interviewed for the study. Table 1 describes the characteristics of the survey sample. Of the 146 subjects, 82 (56%) were African American and 64 (44%) were white. African Americans had significantly lower health literacy, lower education, more religious attendance, and poorer health (all with $P < 0.001$).

Table 1 Characteristics of the Study Sample

Characteristic	Total
No. of subjects	146
Age in years, mean (SD)	57 (12)
Women, *n* (%)	93 (64)
Race, *n* (%)	
African American	82 (56)
White	64 (44)
Health literacy, *n* (%)	
Low	27 (18)
Marginal	30 (21)
Adequate	89 (61)
Education, *n* (%)	
High school or less	82 (56)
College or beyond	64 (44)
Marital status, *n* (%)	
Married	79 (54)
Nonmarried	67 (46)
Religion, *n* (%)	
Catholic/Protestant	72 (49)
Christian (other)	46 (32)
Other	28 (19)
Religious attendance, *n* (%)	
Two times per month or more	63 (43)
One time per month or less	23 (16)
Never	60 (41)
Self-reported health status, *n* (%)	
Very healthy	77 (53)
Somewhat healthy	63 (43)
Not healthy	6 (4)

Outcomes

After hearing a brief verbal description of advanced dementia, 102 (69%) subjects preferred comfort care, 23 (16%) desired life-prolonging care, 14 (10%) chose limited care, and 7 (5%) were unsure of their preferences (Figure 1). The subjects' preferences changed significantly after the video: 136 (93%) of the subjects chose comfort care, 1 (1%) desired life-prolonging care, 8 (5%) chose limited care, and, 1 (1%) was unsure ($P < 0.001$).

Before watching the video, the mean score for uncertainty was 12.8 ($s = 2.6$; 95% confidence interval [CI] = 12.4–13.2). After watching the video, the mean uncertainty score improved to 14.3 ($s = 1.7$; 95% CI = 14.0–14.5). The mean change in uncertainty score after viewing the video was 1.5 ($s = 2.4$; 95% CI = 1.1–1.9; $P < 0.0001$).

Health literacy level was associated with uncertainty level before ($P < 0.001$) and after ($P = 0.046$) watching the video and with the change in uncertainty between the 2 time points ($P = 0.002$). Of the 27 subjects with low health literacy, the mean prevideo uncertainty score was 10.8 ($s = 3.0$; 95% CI = 9.6–12.0) and the mean postvideo uncertainty score was 13.6 ($s = 2.8$; 95% CI = 12.5–14.7). The mean change in uncertainty score in this group was 2.8 ($s = 3.5$; 95% CI = 1.4–4.2).

Of the 30 subjects with marginal health literacy, the mean prevideo uncertainty score was 12.4 ($s = 2.2$; 95% CI = 11.6–13.2) and the mean postvideo uncertainty score was 14.1 ($s = 1.7$; 95% CI = 13.4–14.7). The mean change in uncertainty score in this group was 1.7 ($s = 2.3$; 95% CI = 0.8–2.5).

Of the 89 subjects with adequate health literacy, the mean prevideo uncertainty score was 13.5 ($s = 2.2$; 95% CI = 13.1–14.0) and the mean postvideo uncertainty score was 14.5 ($s = 1.1$; 95% CI = 14.3–14.7). The mean change in uncertainty score in this group was 1.0 ($s = 1.9$; 95% CI = 0.6–1.4).

In addition to health literacy, race, education, religious affiliation, and health status were found to be associated with level of uncertainty before or after watching the video. Race and education were highly correlated with health literacy and were dropped from the model. Overall, 88% of Caucasians had adequate health literacy compared with 40% of African Americans; 72% of those with a high school or higher education had adequate health literacy, whereas only 4% of those with less than high

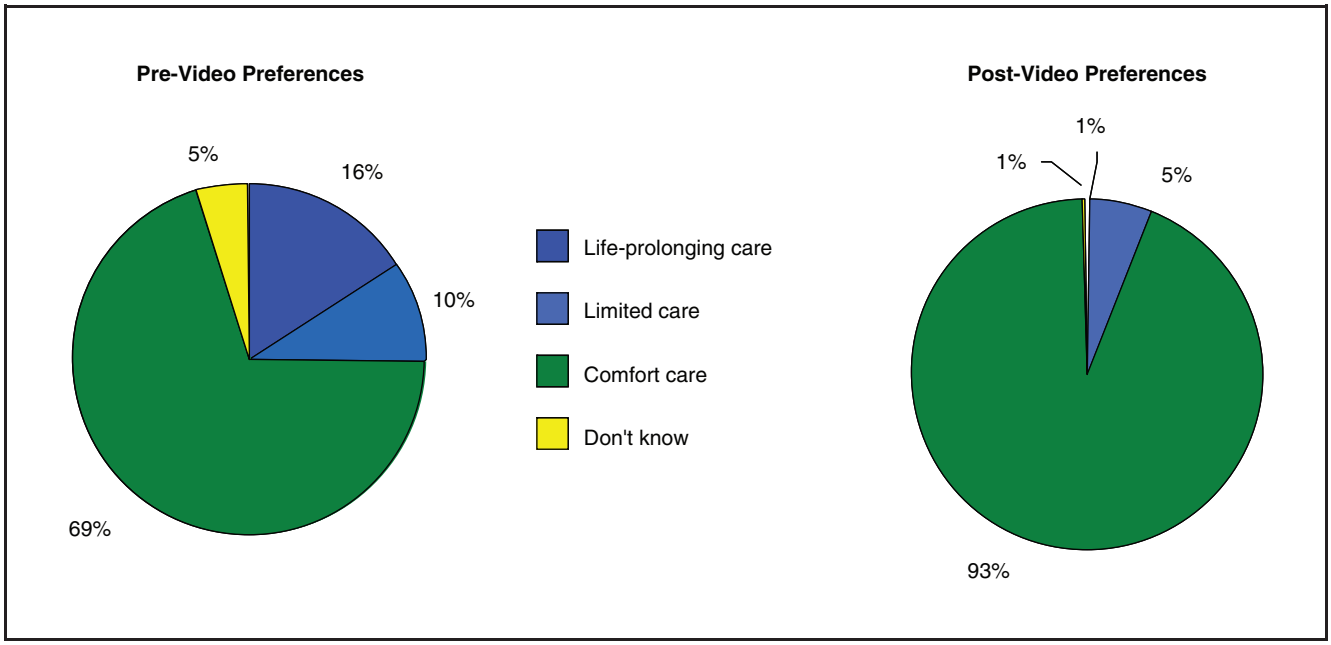


Figure 1 Prevideo and Postvideo preferences.

school education had adequate health literacy (both with $P < 0.001$).

The 3 final multivariable regression models of uncertainty prior to the video, uncertainty after the video, and the change in uncertainty included health literacy, religious affiliation, and health status. The adjusted mean scores from these 3 models are summarized in Table 2. Health literacy remained a significant predictor for the uncertainty prior to the video ($P < 0.001$) and the change in uncertainty ($P = 0.023$) but was no longer statistically significant in the prediction of uncertainty after the video ($P = 0.22$).

DISCUSSION

Multiple studies have shown that decision aids improve the quality of subjects' decisions, including video decision aids at the end of life.[12] Our study suggests that patients with limited health literacy—a significant portion of our patient population—have more uncertainty regarding their decisions at the end of life than those with adequate health literacy and benefit from the use of video decision aids at the end of life more than patients with adequate health literacy. Indeed, after the video, health literacy was no longer independently associated with uncertainty. Decreasing uncertainty remains a significant sphere in which to improve decision making for patients.

The observed changes in overall uncertainty scores and uncertainty scores for the low-literacy group after the video are likely to be clinically important. As others have noted,[16] changes in health-related quality-of-life measurements that exceed 50% of the baseline standard deviation can be considered clinically significant. The changes observed in our study overall and especially for the low-literacy group were well above this cutoff (overall change in uncertainty score divided by standard deviation at baseline = 1.5/2.6 = 0.58; change in uncertainty score in the low-literacy group divided by the standard deviation at baseline = 2.8/3.0 = 0.93). As we hypothesized, the clinical importance of the observed changes was greatest in the group with low health literacy.

The Cochrane database[17] includes multiple decision aids in a variety of clinical contexts in which uncertainty is improved. To our knowledge, this is the 1st study to examine the influence of health literacy on the uncertainty of end-of-life decision making. Our study needs to be interpreted in the context of end-of-life decision making and the study design.

Video clips can be manipulated to favor a particular perspective. To minimize the introduction of bias in the filming of the patient, the video was filmed and edited with close collaboration from expert geriatricians and neurologists. All filming and editing was done by the principal investigator in the cinema verité style of documentary film

Table 2 Adjusted Mean Uncertainty Scores for Prevideo, Postvideo, and Change in Level of Uncertainty by Level of Health Literacy (Adjusted for Health Status and Religious Affiliation)

		Prevideo	Postvideo	Change
Health Literacy	**Subject (*n*)**	**Mean (95% CI)**	**Mean (95% CI)**	**Mean (95% CI)**
Low	27	10.3 (9.3–11.4)	13.9 (13.1–14.6)	3.5 (2.5–4.6)
Marginal	30	11.9 (10.8–12.9)	14.4 (13.6–15.1)	2.5 (1.5–3.5)
Adequate	89	12.5 (11.7–13.3)	14.6 (14.0–15.1)	2.0 (1.3–2.8)
P		< 0.001	0.22	0.023

Note: CI = confidence interval.

making, avoiding the use of special effects or staging. Nonetheless, the use of video inherently involves aesthetic biases, and we have addressed the controversies involving visual media elsewhere.[18] We also did not test the robustness of our findings using other video clips varying features of the patient, such as race/ethnicity, gender, and nursing home setting.

We used a before-and-after study design in which individual subjects heard the description of dementia twice, verbally and then with the video. Subjects may have benefited merely from the repetition of the information for comprehension. It is possible that subjects with limited literacy had more uncertainty in this study because they had less knowledge about dementia. We did not measure knowledge of dementia or goals of care options. Furthermore, although we did conduct multivariable analyses that exhibited the independent predictive value of limited literacy for uncertainty prior to the video, we were unable to fully evaluate the relationship between literacy, education, and religion. Future randomized studies can further evaluate these issues, manage unmeasured confounders, and isolate the effect of the video.

Our study group also did not include the largest minority group in this country, namely, Latinos. Studying the use of video in other languages and other large minority groups would be helpful. Proceeding with such research, however, will require further validation of tools to measure uncertainty and literacy.

We did not analyze subjects' uncertainty in relation to their preferences for end-of-life care nor examine level of uncertainty in those subjects who changed preferences after the video. Such analyses require large numbers of subjects, and research on the relationship between stability of preferences and uncertainty would be interesting. Similarly, although the overall change in uncertainty represented more than a half standard deviation, it is important to note that the true clinical significance of even this large change in uncertainty has not been elucidated.

Our present study asked questions about uncertainty relating to advanced dementia, a common end-of-life scenario. Uncertainty for end-of-life care in other disease states may be different; however, we suspect this to be unlikely. If our results are found to be generalizable to other health states and other populations, this would have important implications for research and practice relating to decision making.

Our data support the idea that the level of uncertainty is more like a state than a trait; that is, people with limited health literacy do not appear to be constitutionally less certain. To the extent that uncertainty may lead to more passivity overall regarding health promotion and self-care activities, our observations may represent an important direction for future research, as the level of uncertainty has not yet been explored as a mechanism for how limited health literacy leads to worse health outcomes. Research evaluating these connections may identify important opportunities for intervention.

Video decision aids are widely accessible and easily disseminated. Research relating to end-of-life decision making may need to focus on decision aids that are proven to benefit patients with limited health literacy. Research of this kind has the potential to improve the quality of decision making for patients.

APPENDIX

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.

REFERENCES

1. Tulsky JA. Beyond advance directives: importance of communication skills at the end of life. JAMA. 2005;294:359–65.

2. Emanuel LL, Danis M, Pearlman RA, Singer PA. Advance care planning as a process: structuring the discussions in practice. J Am Geriatr Soc. 1995;43:440–6.

3. Gillick MR. Advance care planning. N Engl J Med. 2004;350: 7–8.

4. Koch T. Future states: the axioms underlying prospective, future-oriented, health planning instruments. Soc Sci Med. 2001; 52:453–65.

5. Volandes A, Paasche-Orlow M, Gillick MR, et al. Health literacy not race predicts end-of-life preferences. J Gen Intern Med. 2007;22:5.

6. US Department of Education, IoES, National Center for Education Statistics. 2003 National Assessment of Adult Literacy. Available from: URL: http://nces.ed.gov/NAAL/index.asp?file = AssessmentOf/HealthLiteracy.asp&PageID = 161. Accessed 1 April 2007.

7. 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.

8. Frosch DL, Kaplan RM, Felitti VJ. A randomized controlled trial comparing Internet and video to facilitate patient education for men considering the prostate specific antigen test. J Gen Intern Med. 2003;18:781–7.

9. Morgan MW, Deber RB, Llewellyn-Thomas HA, et al. Randomized, controlled trial of an interactive videodisc decision aid for patients with ischemic heart disease. J Gen Intern Med. 2000;15: 685–93.

10. Edwards A, Elwyn G, Mulley A. Explaining risks: turning numerical data into meaningful pictures. BMJ. 2002;324:827–30.

11. Houts PS, Doak CC, Doak LG, Loscalzo MJ. The role of pictures in improving health communication: a review of research on attention, comprehension, recall, and adherence. Patient Educ Couns. 2006;61:173–90.

12. Volandes AE, Lehmann LS, Cook EF, Shaykevich S, Abbo ED, Gillick MR. Using video images of dementia in advance care planning. Arch Intern Med. 2007;167:828–33.

13. Sclan SG, Reisberg B. Functional assessment staging (FAST) in Alzheimer's disease: reliability, validity, and ordinality. Int Psychogeriatr. 1992;4:55–69.

14. Doak CC, Doak LG, Root JH. Teaching Patients with Low Literacy Skills. 2nd ed. Philadelphia: J.B. Lippincott; 1996.

15. O'Connor AM. Validation of a decisional conflict scale. Med Decis Making. 1995;15:25–30.

16. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. Med Care. 2003;41:582–92.

17. O'Connor AM, Stacey D, Rovner D, et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev. 2001:CD001431.

18. Gillick MR, Volandes AE. The psychology of using and creating video decision aids for advance care planning. In: Lynch TE, ed. Psychology of Decision Making in Medicine and Health Care. New York: Nova Science Publishers; 2007. p 193–206.

Physical Health and Drinking Among Medical Inpatients With Unhealthy Alcohol Use: A Prospective Study[1]

Emily C. Williams, Tibor Palfai, Debbie M. Cheng, Jeffrey H. Samet, Katharine A. Bradley, Thomas D. Koepsell, Thomas M. Wickizer, Patrick J. Heagerty, and Richard Saitz

Objective: Unhealthy alcohol use is common in medical inpatients, and hospitalization has been hypothesized to serve as a "teachable moment" that could motivate patients to decrease drinking, but studies of hospital-based brief interventions have often not found decreases. Evaluating associations between physical health and subsequent drinking among medical inpatients with unhealthy alcohol use could inform refinement of hospital-based brief interventions by identifying an important foundation on which to build them. We tested associations between poor physical health and drinking after hospitalization and whether associations varied by alcohol dependence status and readiness to change.

Methods: Participants were medical inpatients who screened positive for unhealthy alcohol use and consented to participate in a randomized trial of brief intervention (n = 341). Five measures of physical health were independent variables. Outcomes were abstinence and the number of heavy drinking days (HDDs) reported in the 30 days prior to interviews 3 months after hospitalization. Separate regression models were fit to evaluate each independent variable controlling for age, gender, randomization group, and baseline alcohol use. Interactions between each independent variable and alcohol dependence and readiness to change were tested. Stratified models were fit when significant interactions were identified.

Results: Among all participants, measures of physical health were not significantly associated with either abstinence or number of HDDs at 3 months. Having an alcohol-attributable principal admitting diagnosis was significantly associated with fewer HDDs in patients who were nondependent [adjusted incidence rate ratio (aIRR) 0.10, 95% CI 0.03–0.32] or who had low alcohol problem perception (aIRR 0.36, 95% CI 0.13–0.99) at hospital admission. No significant association between alcohol-attributable principal admitting diagnosis and number of HDDs was identified for participants with alcohol dependence or high problem perception.

Conclusions: Among medical inpatients with nondependent unhealthy alcohol use and those who do not view their drinking as problematic, alcohol-attributable illness may catalyze decreased drinking. Brief interventions that highlight alcohol-related illness might be more successful.

Key Words: Hospital-Based Brief Intervention, Physical Health, Readiness to Change, Unhealthy Alcohol Use, Alcohol Dependence.

MEDICAL ILLNESS AND poor physical health status are common among patients with unhealthy alcohol use (Blow et al., 2000; Bridevaux et al., 2004; Chou et al., 1996; Green et al., 2004; Rehm et al., 2003; Room et al., 2005; Saitz et al., 2006; Solberg et al., 2008; Stein, 1999; Williams et al., 2010), a term which is used to describe a spectrum from drinking amounts that risk health consequences (e.g., above recommended limits), to drinking amounts

From the VA Health Services Research & Development (ECW, KAB), Seattle, WA; Departments of Medicine (KAB), Health Services (ECW, KAB, TDK, TMW), Epidemiology (TDK), and Biostatistics (PJH), University of Washington, Seattle, WA; Clinical Addiction Research and Education (CARE) Unit, Section of General Internal Medicine, Boston Medical Center and Boston University School of Medicine (TP, DMC, JHS, RS), Boston, MA; Department of Psychology (TP), Boston University, Boston, MA; and Departments of Biostatistics (DMC), Epidemiology (RS), and Social & Behavioral Sciences (JHS), Boston University School of Public Health, Boston, MA.

Received for publication November 3, 2009; accepted February 17, 2010.

[1]Views expressed in this article are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs, the University of Washington, or Boston University. Preliminary versions of this work were presented at the 2007 meeting of the International Network on Brief Interventions for Alcohol Problems and the 2008 Research Society on Alcoholism annual meeting, as well as included in the dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy from the University of Washington (ECW). This study was supported by NIAAA grant R01-AA 12617 (Saitz, PI), and the material presented here is the result of work supported by resources from the VA Puget Sound Health Care System, Seattle, Washington.

Reprint requests: Emily C. Williams, PhD, MPH, VA Puget Sound Health Care System, 1100 Olive Way, Suite 1400, Seattle, WA 98101; Fax: +1-206-764-2935; E-mail: emily.williams3@va.gov

DOI: 10.1111/j.1530-0277.2010.01203.x

associated with consequences but not yet meeting criteria for a disorder, to meeting diagnostic criteria for alcohol abuse or dependence (Saitz, 2005). Approximately 25% of all persons admitted to general hospitals have alcohol use disorders or are being treated for the consequences of their drinking, making hospitalization a potentially opportune time for interventions to reduce unhealthy alcohol use (Smothers et al., 2003). However, the evidence in support of brief counseling interventions in hospital settings is mixed, often showing lack of efficacy (Chick et al., 1985; Elvy et al., 1988; Emmen et al., 2004; Freyer-Adam et al., 2008; Heather et al., 1996; Holloway et al., 2007; Persson and Magnusson, 1989; Saitz et al., 2007, 2009). This lack of efficacy may be due in part to the high prevalence of dependence among patients with unhealthy alcohol use in hospitals (Saitz et al., 2006), although studies that excluded dependent patients have also had mixed results (McQueen et al., 2009). Improvements in opportunistic counseling interventions for hospitalized patients identified with unhealthy use, including dependence, are sorely needed.

Qualitative studies have identified factors responsible for catalyzing change in patients with unhealthy alcohol use (Orford and Hawker, 1974; Orford et al., 2006, 2008), which include "awareness of accumulating harms" and "triggering occurrences" (Orford et al., 2006). As such, poor physical health may serve as a catalyst of change. Further, in the context of a hospitalization, which has been hypothesized to provide a "teachable moment," patients with unhealthy alcohol use may recognize the link between their drinking and health and be particularly amenable to making changes in their drinking (Figlie et al., 2005; Stewart and Connors, 2007), with or without intervention. As such, patients with the physical health consequences from alcohol use would be more likely to change than those without such consequences.

Factors that may catalyze change among patients with unhealthy alcohol use may operate differently for patients based on both the severity of dependence symptoms and their readiness for or commitment to change. Specifically, patients with alcohol dependence, who typically exhibit impaired control over their drinking, or those who are not ready to change may be less likely to decrease drinking after a catalyst than those without dependence or those with greater readiness to change (DiMartini et al., 2002; Kelly et al., 2006; Moos and Moos, 2006; Orford et al., 2008; Saitz et al., 2009; Vielva and Iraurgi, 2001; Walton et al., 2003).

Evaluating associations between physical health and subsequent drinking among medical inpatients with unhealthy alcohol use could help to inform further refinement of hospital-based brief interventions by identifying whether they should specifically focus on physical health. This study sought, in a secondary analysis of data collected prospectively, to evaluate whether, independent of an intervention, 5 different measures of physical health were associated with two drinking outcomes 3 months after hospitalization and to explore whether associations between physical health and drinking vary by alcohol dependence or readiness to change.

METHODS

Study Design

We studied a prospective cohort of adult medical inpatients with unhealthy alcohol use who were first identified by screening in the hospital and then enrolled in a randomized trial of brief alcohol counseling. Although the intervention was not associated with decreased drinking or receipt of alcohol-related treatment in this trial (Saitz et al., 2007), some subgroups, including patients without alcohol dependence, appeared to benefit from the brief intervention (Saitz et al., 2009). The study represents a secondary analysis of data collected during this trial and was approved by the institutional review boards of Boston Medical Center and the University of Washington. Subjects provided informed consent to participate in the trial and were compensated for each completed interview. A federal government certificate of confidentiality was obtained to protect participant privacy.

Participants

Participants included 341 adult medical inpatients who reported drinking risky amounts in the past month defined as >14 standard drinks per week or ≥ 5 drinks per occasion for men (>11 and ≥ 4, respectively, for both women and people ≥ 66 years (i.e., screened positive for unhealthy alcohol use) based on the screening strategy recommended by the National Institute on Alcohol Abuse and Alcoholism (National Institute on Alcohol Abuse and Alcoholism et al., 2007). Participants also were fluent in English or Spanish, provided names of 2 contacts to assist with follow-up, had no plans to move away from the study area in the following year, scored 21 or more points on the Mini-Mental State Examination (Smith et al., 2006), and consented to participate in the trial (Saitz et al., 2007).

Assessments

Participants were assessed via in-person interviews by trained research associates at enrollment and a 3-month follow-up visit. Also at enrollment, a physician-researcher reviewed each participant's medical record to determine medical diagnoses, including the principal reason for admission (principal admitting diagnosis).

Measures

Drinking Outcomes. Alcohol use in the 30 days before each interview was assessed using the validated Timeline Follow-Back calendar method (Sobell et al., 1988), which identifies the number of standard drinks consumed on each of the past 30 days. Two drinking measures derived from the Timeline Follow-Back were used as study outcomes: (i) 30-day abstinence and (ii) number of heavy drinking days (HDDs) reported in the 30 days prior to the 3-month assessment. HDDs were defined as drinking ≥ 5 drinks per day for men and ≥ 4 per day for women.

Independent Variables. Five measures of different dimensions of physical health, the independent variables for the study, were derived from interview and structured record review at the time of enrollment: (i) recent medical comorbidity, (ii) lifetime medical comorbidity, (iii) self-reported physical health status, (iv) any alcohol-attributable medical diagnosis, and (v) alcohol-attributable principal admitting diagnosis. Medical comorbidities were assessed using a questionnaire validated by Katz, et al., which is similar to the Charlson Comorbidity Index (Charlson et al., 1994) but is based on patient report instead of administrative data and assesses the presence of medical illnesses including kidney disease, diabetes, heart disease, chronic obstructive pulmonary disease, AIDS, and others using both recent (past 3 months) and lifetime timeframes (Katz et al., 1996). We categorized the number of recent and lifetime comorbidities (0, 1, or ≥ 2). The

physical component summary (PCS) score from the 12-item Short Form Health Survey (Ware et al., 1996) measured physical health status. The PCS score ranges from 1 to 100 and is standardized to the U.S. population mean with 1 representing poor and 100 representing perfect physical health status (Ware et al., 1996). To avoid an assumption of linearity, PCS scores were categorized into quintiles with the highest quintile (best health) used as the referent category. Finally, a physician-researcher reviewed each patient's medical record at enrollment for any medical diagnoses that are 100% alcohol-attributable, and for whether any of these were the principal admitting diagnosis. Diagnoses included alcohol intoxication, alcoholic pellagra, alcoholic polyneuropathy, alcoholic gastritis, alcohol cardiomyopathy, alcoholic liver damage, acute alcoholic cirrhosis of the liver, alcoholic hepatitis, alcoholic fatty liver, alcoholic amnestic syndrome, other alcoholic dementia, alcohol withdrawal, alcohol withdrawal hallucinosis, alcohol withdrawal delirium, other alcoholic psychosis, alcohol withdrawal convulsion, or alcoholic pancreatitis (Adams et al., 1993).

Potential Effect Modifiers. DSM IV diagnosis of current alcohol dependence (American Psychiatric Association, 1994) was made at enrollment based on the Composite International Diagnostic Interview (CIDI) Alcohol Module (Robins et al., 1988). Readiness to change drinking was assessed via interview with the validated Stages of Change Readiness and Treatment Eagerness Scale (SOCRATES), which was developed to assess readiness among individuals presenting for specialty alcohol treatment (Miller and Tonigan, 1996). Two SOCRATES components (*Problem Perception* and *Taking Action*) were used rather than the 3 factors originally described based on results from a previous factor analysis in this patient population (Bertholet et al., 2009b), because it is consistent with another prior study (Maisto et al., 1999) and because these factors more aptly describe change readiness among patients not actively seeking or receiving specialty alcohol treatment. The two components yield continuous scores (range: 10–50 for Problem Perception, 6–30 for Taking Action) with higher scores representing more readiness. The readiness scores have been previously modeled using quartiles of the distribution because of nonlinear associations with drinking outcomes (Bertholet et al., 2009b). For this study, each score was dichotomized based on the observed associations with the outcomes of interest for this study and for ease of interpretation. Based on results of preliminary analyses categorizing readiness scores using quartiles of the distribution, Problem Perception was dichotomized at the 50th and Taking Action at the 75th percentile; residual plots and likelihood ratio tests evaluating goodness-of-fit suggested adequate fit.

Covariates. Covariates representing potential confounders were selected based on known associations between medical illness or health status and alcohol use and included: demographic characteristics (age in years, female gender, and race including Hispanic, Black, White, and other); smoking status [never, current, and past (Fagerstrom Tolerance Test), and quantity/frequency questions] (Pomerleau et al., 1994), and alcohol use reported at enrollment. Number of drinks per day (continuous variable) was included in models with abstinence as the outcome, and the number of heavy drinking days in the 30 days prior to baseline (categories: 0–1, 2–14, and ≥15 heavy drinking days) was included in models with number of heavy drinking days as the outcome. For evaluating the association between measures of physical health and abstinence, alcohol dependence was considered a potential confounder (in addition to a potential effect modifier) because of its strong associations with both medical illness and abstinence (Dawson et al., 2005). Despite the negative main trial results, randomization group was also included as a covariate because we sought to identify associations independent of intervention, subgroup analyses suggested some intervention effect, and the intervention may have resulted in differences in drinking or readiness to change between groups (Saitz et al., 2009).

Analyses

Subjects lost to follow-up were compared with those assessed at 3 months on all patient demographic and drinking characteristics using chi-square statistics, two-sample *t*-tests, and nonparametric (two-sample Wilcoxon rank-sum or Mann–Whitney) tests as appropriate. Separate logistic regression models were fit to evaluate the association between each of the 5 independent variables and 30-day abstinence. Negative binomial regression was used to model the association between each of the independent variables and the number of heavy drinking days reported at 3 months, because days are counts with a skewed distribution (large numbers of zeros from abstinent subjects, and some extreme numbers of days). Negative binomial regression generates estimates of the incidence rate ratio (IRR), which is interpreted as a multiplicative increase or decrease in the number of days for each 1-unit difference in the predictor variable (similar to a relative risk). A robust variance correction was used to allow for over-dispersion. Preliminary unadjusted models were fit first and then adjusted models controlling for all covariates were used as the primary analysis. Multiplicative first-order interactions between each measure of physical health and potential effect modifiers (dependence status and readiness to change) were tested in both unadjusted and adjusted models. Stratified analyses were conducted when interactions were significant ($p < 0.05$). We did not formally adjust for multiple comparisons because this secondary analysis was exploratory and intended to assess whether a single concept—physical health—was associated with drinking 3 months after hospitalization. All analyses were performed using Stata Version 10.1 (StataCorp., 2007) .

RESULTS

Participants had a mean age of 44 years and were racially diverse (45% Black, 39% White, and 9% Hispanic) (Table 1). Only 14% of participants reported never smoking; 75% reported current smoking. At baseline, patients reported consuming a median of 3.6 drinks per day, 38% of patients reported 15 or more heavy drinking days, 4% met diagnostic criteria for current alcohol abuse, and 77% met criteria for current alcohol dependence. Patients generally reported high levels of recognition of unhealthy drinking and efforts to change. Mean problem perception score was 35 (SD 11, range 10–50), and mean Taking Action score was 21 (SD 6, range 6–30). Participants had poor average physical health status [Physical Component Summary score of 38 (SD 9)]; 22% reported 2 or more recent medical comorbidities, and 33% reported 2 or more lifetime comorbidities. Forty-six percent of participants had any alcohol-attributable medical diagnosis in their medical record, while 15% had an alcohol-attributable principal admitting diagnosis.

Overall, 272 patients were interviewed 3 months after hospitalization (80% follow-up). No statistically significant differences (at $p < 0.05$) in demographic or drinking characteristics were identified between those with and without follow-up. However, a nonsignificantly greater proportion of patients with follow-up, compared to those without, met diagnostic criteria for alcohol dependence (79% vs. 68%,

Table 1. Participant Characteristics at Baseline (n = 341)

	Mean (SD)
Age	44 (11)
	N (%)
Female	99 (29)
Race	
Hispanic	30 (9)
Black	155 (45)
White	133 (39)
Other	23 (7)
Smoking Status	
Never	48 (14)
Current	257 (75)
Past	36 (11)
Heavy Drinking Days (in the prior 30 days)	
0–1	59 (17)
2–14	152 (45)
≥15	130 (38)
Median Drinks per Day	3.6
Alcohol Dependent	261 (77)
Alcohol Abuse	15 (4)
	Mean (SD)
Problem Perception (Range 10–50)	35 (11)
Low, Score ≤38 (n = 172)	26 (9)
High, Score >38 (n = 166)	44 (4)
Taking Action (Range 6–30)	21 (6)
Low, Score ≤25 (n = 255)	19 (5)
High, Score >25 (n = 85)	28 (2)
	Mean (SD)
Physical Health Status (PCS Score; Range 17.8–61.9)	38 (9)
(Worst Health) Quintile 1 (17.8–30.3)	26 (3)
Quintile 2 (30.4–34.5)	33 (1)
Quintile 3 (34.6–39.4)	37 (1)
Quintile 4 (39.5–36.0)	43 (2)
(Best Health) Quintile 5 (46.2–61.9)	52 (4)
	N (%)
Recent # Medical Comorbidities	
0	148 (48)
1	96 (31)
≥2	67 (22)
Lifetime # Medical Comorbidities	
0	107 (34)
1	102 (33)
≥2	104 (33)
Any Alcohol-Attributable Medical Diagnosis	156 (46)
Alcohol-Attributable Principal Admitting Diagnosis	51 (15)

p = 0.06). There were no significant differences in measures of physical health between those who did and did not complete follow-up.

At 3 months, 60 participants (22% of those with follow-up) were abstinent, and the median number of heavy drinking days was 5 (range 0–30). Most measures of physical health were not significantly associated with either abstinence or the number of heavy drinking days reported at 3 months in both unadjusted and adjusted regression models (Table 2). However, in unadjusted models, patients in the lowest quintile of self-reported physical health status (worst health) had significantly greater odds of abstinence at 3 months compared to those in the highest quintile (OR 3.35, 95% CI 1.25–8.96). The significance of the association was attenuated after adjustment (Table 2).

Table 3 displays p values for all tests of effect modification. Alcohol dependence significantly modified the association between having any alcohol-attributable medical diagnosis and abstinence and between having an alcohol-attributable principal admitting diagnosis and the number of heavy drinking days reported at 3 months but did not modify any other associations between measures of physical health and either drinking outcome. For the 58 participants without alcohol dependence, having an alcohol-attributable medical diagnosis (n = 13) was significantly associated with increased abstinence in unadjusted analyses (OR 4.36; 95% CI: 1.12–17.03) but not after adjustment for potential confounders (OR 3.83; 95% CI 0.75–19.59). For those who were alcohol dependent (n = 214), there was no significant association between having an alcohol-attributable diagnosis (n = 114) and abstinence in unadjusted or adjusted analyses (ORs 0.90; 95% CI 0.47–1.70 and 0.78; 95% CI 0.41–1.62, respectively). For the 58 participants who were not alcohol dependent, having an alcohol-attributable principal admitting diagnosis (n = 4) was significantly associated with fewer heavy drinking days in unadjusted and adjusted models (Table 4). For the 214 participants who were alcohol dependent, there was no significant association between having an alcohol-attributable principal admitting diagnosis (n = 37) and number of heavy drinking days (Table 4).

Neither measure of readiness to change modified any of the associations between measures of physical health and abstinence in unadjusted or adjusted analyses (Table 3). In contrast, both measures of readiness to change modified associations between having an alcohol-attributable principal admitting diagnosis and the number of heavy drinking days at 3 months (Table 3). Associations in unadjusted and adjusted analyses stratified by levels of Taking Action were not significant. However, among subjects with low Problem Perception (n = 132), those with an alcohol-attributable principal admitting diagnosis (n = 9) had 0.36 times the number of heavy drinking days compared to those without such a diagnosis (95% CI 0.13–0.99) (Table 4). No association was detected between having an alcohol-attributable principal admitting diagnosis (n = 30) and heavy drinking days among participants with high Problem Perception (n = 137). Problem Perception also modified the association between self-reported physical health status and heavy drinking days (Table 3). Among the 132 patients with low Problem Perception, those in the second Quintile of health status (n = 20) had 2.06 times the number of heavy drinking days (95% CI 1.17–3.62) compared to those in the referent group with the highest health status [5th Quintile (n = 33)] in unadjusted analyses. However, associations in stratified adjusted analyses were all not significant. No other significant interactions were found between readiness to change and the independent variables in models of heavy drinking days at 3 months (Table 3).

Table 2. Unadjusted and Adjusted Associations Between Measures of Physical Health and 3-Month Drinking Outcomes

	30 Day Abstinence[a]*				Number of Heavy Drinking Days[b]*			
	Unadjusted Models		Adjusted Models		Unadjusted Models		Adjusted Models	
	OR	95% CI	OR	95% CI	IRR	95% CI	IRR	95% CI
Recent # of Medical Comorbidities								
0	–	–	–	–	–	–	–	–
1	1.60	0.78–3.29	1.69	0.80–3.59	1.04	0.75–1.43	0.93	0.66–1.31
≥2	1.51	0.67–3.39	1.44	0.62–3.34	1.34	0.90–1.98	1.18	0.78–1.79
Lifetime # of Medical Comorbidities								
0	–	–	–	–	–	–	–	–
1	1.04	0.48–2.25	1.13	0.51–2.51	1.03	0.75–1.40	0.95	0.68–1.32
≥2	1.26	0.59–2.70	1.15	0.51–2.58	1.22	0.84–1.77	1.10	0.73–1.64
Physical Health Status								
Quintile 1	3.35	1.25–8.96	2.81	0.98–8.00	1.07	0.66–1.71	0.96	0.60–1.54
Quintile 2	1.34	0.47–3.82	1.00	0.32–3.03	1.24	0.81–1.89	1.18	0.78–1.80
Quintile 3	1.54	0.54–4.33	1.10	0.36–3.38	1.09	0.72–1.65	1.04	0.67–1.61
Quintile 4	1.95	0.71–5.37	1.74	0.61–4.95	1.00	0.66–1.52	0.96	0.63–1.45
Quintile 5	–	–	–	–	–	–	–	–
Any Alcohol-Attributable Medical Diagnosis	1.26	0.71–2.23	1.13	0.60–2.10	0.90	0.66–1.22	0.83	0.60–1.14
Alcohol-Attributable Principal Admitting Diagnosis	1.81	0.87–3.76	1.87	0.85–4.11	0.97	0.65–1.44	0.88	0.58–1.33

[a]Adjusted models included age, gender, race, smoking status, randomization group, alcohol dependence, and drinks/day at baseline.
[b]All models were adjusted for heaving drinking days at baseline; adjusted models also include age, gender, race, smoking status, randomization group, and alcohol dependence.
*p-Values for all associations were all >0.05. p-Values obtained for categorical independent variables were the result of tests of trend; p-values obtained for dichotomous independent variables were the results of Wald tests.

Table 3. Results of Tests of Effect Modification: p-Values for Interactions Between Measures of Physical Health and Alcohol Dependence and Readiness to Change (Significant Results are Bolded)

	30-Day Abstinence[b]			# Heavy Drinking Days[c]		
Independent Variables[a]	Dependent	Problem Perception	Taking Action	Dependent	Problem Perception	Taking Action
Recent # of Medical Comorbidities						
Unadjusted	0.67	0.73	0.56	0.76	0.52	0.16
Adjusted	0.71	0.74	0.66	0.60	0.62	0.12
Lifetime # of Medical Comorbidities						
Unadjusted	0.19	0.90	0.90	0.09	0.75	0.95
Adjusted	0.25	0.85	0.80	0.08	0.72	0.98
Physical Health Status						
Unadjusted	0.83	0.27	0.12	0.28	**0.02**	0.87
Adjusted	0.87	0.21	0.09	0.21	**0.01**	0.85
Any Alcohol-Attributable Medical Diagnosis						
Unadjusted	**0.04**	0.80	0.18	0.12	0.17	0.75
Adjusted	0.05	0.83	0.08	0.12	0.22	0.91
Alcohol-Attributable Principal Admission						
Unadjusted	0.94	0.70	0.18	**0.04**	0.07	0.13
Adjusted	0.92	0.70	0.06	**0.02**	**0.04**	**0.03**

[a]Medical comorbidities (both recent and lifetime) modeled categorically (0, 1, ≥2) with 0 the referent; Physical Health Status modeled categorically in Quintiles (Quintile 5 referent); Alcohol-Attributable diagnoses (both any and principal admission) modeled dichotomously.
[b]Adjusted models included age, gender, race, smoking status, randomization group, alcohol dependence, and drinks per day at baseline.
[c]All models were adjusted for heaving drinking days at baseline; adjusted models also include age, gender, race, smoking status, randomization group, and alcohol dependence.

DISCUSSION

In this study of medical inpatients with unhealthy alcohol use, most participants had alcohol dependence and reported substantial medical comorbidity. Among all participants, there were no significant associations between measures of physical health and either abstinence or heavy drinking days 3 months after hospitalization, though we did observe a borderline significant association between worse physical health status and abstinence. However, among participants with nondependent unhealthy alcohol use and those who were less aware of their drinking as a problem regardless of dependence status, having an alcohol-attributable principal admitting diagnosis was associated with less heavy episodic drinking at follow-up.

Table 4. Associations Between Alcohol-Attributable Principal Admitting Diagnosis and Number of Heavy Drinking Days Stratified by Alcohol Dependence and Problem Perception (Significant Results are Bolded)

	Nondependent (*N* = 58)	Dependent (*N* = 214)	Low Problem Perception (*N* = 132)	High Problem Perception (*N* = 137)
	N (%)	*N* (%)	*N* (%)	*N* (%)
Alcohol-Attributable Principal Admitting Diagnosis	4 (7)	37 (17)	9 (7)	30 (22)
	IRR (95% CI)	IRR (95% CI)	IRR (95% CI)	IRR (95% CI)
Unadjusted[a]	**0.15 (0.04–0.56)**	1.04 (0.71–1.53)	**0.46 (0.21–1.00)**	1.10 (0.83–1.67)
Adjusted[b]	**0.10 (0.03–0.32)**	0.94 (0.62–1.42)	**0.36 (0.13–0.99)**	1.02 (0.61–1.71)

[a]Adjusted for heaving drinking days at baseline.
[b]Adjusted for age, gender, race, smoking status, randomization group, heaving drinking days at baseline, and alcohol dependence.

Several previous studies have explored the idea that medical illness or one's perception of their physical health may be associated with or lead to changes in drinking. Multiple successful alcohol-related interventions have focused on addressing a medical issue via ongoing monitoring of abnormal alcohol-related lab tests or blood pressure (Fleming et al., 2004; Kristenson et al., 1983, 2002; Maheswaran et al., 1992; Willenbring and Olson, 1999). One study of patients with HIV and past alcohol problems found that those who were told they had hepatitis C were more likely to reduce drinking at follow-up than those who were not (Tsui et al., 2007), and interventions in patients identified as heavy drinkers based on physical health problems were associated with reduced drinking and improved health outcomes (Israel et al., 1996; Kristenson et al., 1983; Willenbring and Olson, 1999). Finally, three randomized trials of interventions for unhealthy alcohol use found that the intervention was associated with improved outcomes for patients with an alcohol-related diagnosis but not for those without (Saitz et al., 2009; Walton et al., 2008; Weisner et al., 2001). Only one previous study that examined physical health and alcohol outcomes was conducted in medical inpatients with unhealthy alcohol use (Stewart and Connors, 2007). That study identified associations between readiness to change drinking and self-reported health status and alcohol-related physical consequences but found that only alcohol-related physical consequences were significantly associated with taking steps to change drinking (Stewart and Connors, 2007). Together, findings of previous studies suggest that poor physical health may motivate patients to consider changes in drinking and that, in combination with an intervention, alcohol-related illness can be a strong motivator of actual changes in drinking.

Our findings support and extend findings of previous studies in two ways. First, while the previous study in medical inpatients evaluated associations between self-reported physical health and readiness to change drinking (Stewart and Connors, 2007), our study evaluated associations between both self-report and medical-record-derived measures of physical health and patients' self-reported drinking 3 months after hospitalization. Though not statistically significant, estimates of associations between physical health and drinking went in the expected direction such that poorer physical health was associated with decreased drinking at follow-up. Second, we found that, independent of an intervention, admission to a hospital for a problem resulting from drinking was associated with less heavy drinking for nondependent patients and those who did not perceive their excessive drinking to be problematic at the time of their hospitalization.

Several previous studies have demonstrated that high levels of readiness to change when defined as problem recognition are strongly correlated with more severe unhealthy alcohol use (Maisto et al., 2001; Samet and O'Connor, 1998; Williams et al., 2006, 2007). And, although some measures of readiness to change predict decreases in drinking (Heather et al., 1993; Williams et al., 2007), greater Problem Perception has previously been demonstrated to predict *increases* in drinking (Bertholet et al., 2009a). Further, decreasing drinking may be more difficult for and require more intensive interventions with patients with alcohol dependence (Bischof et al., 2008; Fleming and Manwell, 1999). As such, our finding of decreased incidence of heavy drinking days among participants with nondependent unhealthy alcohol use who may have become aware of consequences of their drinking, but not among participants with alcohol dependence, is consistent with prior research.

We expected but did not find that poor physical health would be associated with decreased drinking among all participants. It could be that the relationship between health and drinking is more complex or that the expected associations are weaker than expected and thus were not statistically significant in this relatively small sample. Our finding that having an alcohol-attributable principal admitting diagnosis was associated with decreased drinking in subgroups of patients may have clinical implications. Brief alcohol counseling interventions that have reliably decreased alcohol consumption in primary care patients do not have confirmed efficacy for medical inpatients with unhealthy alcohol use (Emmen et al., 2004; Freyer-Adam et al., 2008; Holloway et al., 2007; Saitz et al., 2007, 2009). Although further research is needed to confirm this, our findings suggest that being hospitalized for

an alcohol-attributable illness may serve as a catalyst of change (Orford et al., 2006) for someone whose drinking or perception of drinking is more malleable to begin with (i.e., patients without dependence or who have little recognition upon hospital admission of their unhealthy drinking). As such, it could be that hospitalizations for alcohol-attributable illness may be a meaningful aspect to focus on during brief interventions for inpatients with less severe unhealthy alcohol use. But, most patients identified by alcohol screening in hospital settings are alcohol dependent (Saitz et al., 2006), a phenomenon which could account for lack of efficacy of brief interventions in this population (Guth et al., 2008). Further, 45 of the 51 (88%) patients who were hospitalized for alcohol-attributable illnesses in this study met criteria for alcohol dependence, and results of this study suggest that alcohol-attributable illness was not a strong catalyst of change for these patients. However, the study did not address whether an *intervention* focused on poor physical health impacted drinking outcomes. It could be that building upon the presence of an alcohol-attributable illness, or focusing on alcohol-related symptoms or problems (such as trauma) in the absence of alcohol-attributable illness, during brief interventions in hospital settings would help catalyze change for all patients along the spectrum of unhealthy alcohol use.

This study is limited in several ways. First, these analyses were conducted in a cohort of patients who consented to participate in a randomized controlled trial, which may limit the generalizability of findings. However, those who enrolled were identified by screening all medical inpatients. In addition, the parent study was not designed to answer the questions posed and, therefore, our ability to detect associations of the observed magnitude was limited both in the overall analyses and particularly in subgroup analyses exploring effect modification. In particular, the group of patients who were not alcohol dependent, for whom changes in drinking may be easier than they are for alcohol-dependent patients, was small, which may have limited our ability to detect important associations (although it was in this group that we did find an association between physical health and drinking). Future research should be pursued to confirm the results of this exploratory study. However, we found that, in exploring whether multiple dimensions of physical health were associated with changes in drinking, point estimates generally went in the expected direction, some results were significant, and consistent patterns emerged in subgroup analyses. Finally, the observational nature of our data makes us unable to determine a causal association between poor physical health and subsequent drinking. However, because it is impossible to randomize patients to medical illnesses or perceptions of their physical health status as poor, prospective data collection and adjusted analyses as in this study may be the best way to answer the research question.

Despite the limitations, some conclusions can be drawn from these findings, which merit further research. Some dimensions of physical health (e.g., medical illness) may not be associated with subsequent drinking among medical inpatients with unhealthy alcohol use. However, among participants with nondependent unhealthy alcohol use and those less aware that they are drinking at unhealthy levels, being admitted to the hospital for an alcohol-attributable diagnosis was associated with less heavy drinking. Although further research should confirm these findings, they suggest that, among medical inpatients with less severe unhealthy alcohol use, medical illness attributable to alcohol use may serve as a catalyst for positive changes in drinking. Hospital clinicians may be uniquely positioned to offer assistance to inpatients with less severe unhealthy alcohol use by helping patients to recognize a link between their drinking and medical diagnoses.

ACKNOWLEDGMENT

This study was supported by NIAAA grant RO1-AA 12617 (Saitz, PI).

REFERENCES

Adams WL, Yuan Z, Barboriak JJ, Rimm AA (1993) Alcohol-related hospitalizations of elderly people. Prevalence and geographic variation in the United States [published erratum appears in *JAMA* 1993 Nov 3;270:2055]. JAMA 270:1222–1225.

American Psychiatric Association (1994) Diagnostic and Statistical Manual of Mental Disorders, 4th Edn. American Psychiatric Association, Washington, DC.

Bertholet N, Cheng DM, Palfai TP, Samet JH, Saitz R (2009a) Does readiness to change predict subsequent alcohol consumption in medical inpatients with unhealthy alcohol use? Addict Behav 34:636–640.

Bertholet N, Dukes K, Horton NJ, Palfai TP, Pedley A, Saitz R (2009b) Factor structure of the SOCRATES questionnaire in hospitalized medical patients. Addict Behav 34:568–572.

Bischof G, Grothues JM, Reinhardt S, Meyer C, John U, Rumpf HJ (2008) Evaluation of a telephone-based stepped care intervention for alcohol-related disorders: a randomized controlled trial. Drug Alcohol Depend 93:244–251.

Blow FC, Walton MA, Barry KL, Coyne JC, Mudd SA, Copeland LA (2000) The relationship between alcohol problems and health functioning of older adults in primary care settings. J Am Geriatr Soc 48:769–774.

Bridevaux IP, Bradley KA, Bryson CL, McDonell MB, Fihn SD (2004) Alcohol screening results in elderly male veterans: association with health status and mortality. J Am Geriatr Soc 52:1510–1517.

Charlson M, Szatrowski TP, Peterson J, Gold J (1994) Validation of a combined comorbidity index. J Clin Epidemiol 47:1245–1251.

Chick J, Lloyd G, Crombie E (1985) Counselling problem drinkers in medical wards: a controlled study. BMJ 290:965–967.

Chou SP, Grant BF, Dawson DA (1996) Medical consequences of alcohol consumption – United States, 1992. Alcohol Clin Exp Res 20:1423–1429.

Dawson DA, Grant BF, Stinson FS, Chou PS, Huang B, Ruan WJ (2005) Recovery from DSM-IV alcohol dependence: United States, 2001–2002. Addiction 100:281–292.

DiMartini A, Weinrieb R, Fireman M (2002) Liver transplantation in patients with alcohol and other substance use disorders. Psychiatr Clin North Am 25:195–209.

Elvy GA, Wells JE, Baird KA (1988) Attempted referral as intervention for problem drinking in the general hospital. Br J Addict 83:83–89.

Emmen MJ, Schippers GM, Bleijenberg G, Wollersheim H (2004) Effectiveness of opportunistic brief interventions for problem drinking in a general hospital setting: systematic review. BMJ 328:318.

Figlie NB, Dunn J, Gomes LC, Turisco J, Paya R, Laranjeira R (2005) Motivation to change drinking behavior: the differences between alcohol users

from an outpatient gastroenterology clinic and a specialist alcohol treatment service. Sao Paulo Med J 123:223–228.

Fleming M, Brown R, Brown D (2004) The efficacy of a brief alcohol intervention combined with %CDT feedback in patients being treated for type 2 diabetes and/or hypertension. J Stud Alcohol 65:631–637.

Fleming MF, Manwell LB (1999) Brief intervention in primary care settings: a primary treatment method for at-risk, problem, and dependent drinkers. Alcohol Res Health 23:128–137.

Freyer-Adam J, Coder B, Baumeister SE, Bischof G, Riedel J, Paatsch K, Wedler B, Rumpf HJ, John U, Hapke U (2008) Brief alcohol intervention for general hospital inpatients: a randomized controlled trial. Drug Alcohol Depend 93:233–243.

Green CA, Perrin NA, Polen MR (2004) Gender differences in the relationships between multiple measures of alcohol consumption and physical and mental health. Alcohol Clin Exp Res 28:754–764.

Guth S, Lindberg SA, Badger GJ, Thomas CS, Rose GL, Helzer JE (2008) Brief intervention in alcohol-dependent versus nondependent individuals. J Stud Alcohol Drugs 69:243–250.

Heather N, Rollnick S, Bell A (1993) Predictive validity of the Readiness to Change Questionnaire. Addiction 88:1667–1677.

Heather N, Rollnick S, Bell A, Richmond R (1996) Effects of brief counselling among male heavy drinkers identified on general hospital wards. Drug Alcohol Rev 15:29–38.

Holloway AS, Watson HE, Arthur AJ, Starr G, McFadyen AK, McIntosh J (2007) The effect of brief interventions on alcohol consumption among heavy drinkers in a general hospital setting. Addiction 102:1762–1770.

Israel Y, Hollander O, Sanchez-Craig M, Booker S, Miller V, Gingrich R, Rankin JG (1996) Screening for problem drinking and counseling by the primary care physician-nurse team. Alcohol Clin Exp Res 20:1443–1450.

Katz JN, Chang LC, Sangha O, Fossel AH, Bates DW (1996) Can comorbidity be measured by questionnaire rather than medical record review? Med Care 34:73–84.

Kelly M, Chick J, Gribble R, Gleeson M, Holton M, Winstanley J, McCaughan GW, Haber PS (2006) Predictors of relapse to harmful alcohol after orthotopic liver transplantation. Alcohol Alcohol 41:278–283.

Kristenson H, Ohlin H, Hulten-Nosslin M, Trell E, Hood B (1983) Identification and intervention of heavy drinking in middle-aged men: results and follow-up of 24–60 months of long-term study with randomized controls. Alcohol Clin Exp Res 7:203–209.

Kristenson H, Osterling A, Nilsson JA, Lindgarde F (2002) Prevention of alcohol-related deaths in middle-aged heavy drinkers. Alcohol Clin Exp Res 26:478–484.

Maheswaran R, Beevers M, Beevers DG (1992) Effectiveness of advice to reduce alcohol consumption in hypertensive patients. Hypertension 19:79–84.

Maisto SA, Conigliaro J, McNeil M, Kraemer K, Conigliaro RL, Kelley ME (2001) Effects of two types of brief intervention and readiness to change on alcohol use in hazardous drinkers. J Stud Alcohol 62:605–614.

Maisto SA, Conligliaro J, McNeil M, Kraemer KL, O'Connor M, Kelley ME (1999) Factor structure of the SOCRATES in a sample of primary care patients. Addict Behav 24:879–892.

McQueen J, Howe TE, Allan L, Mains D (2009) Brief interventions for heavy alcohol users admitted to general hospital wards. Cochrane Database Syst Rev CD005191.

Miller WR, Tonigan JS (1996) Assessing Drinkers Motivation for Change: the stages of Change Readiness and Treatment Eagerness Scale (SOCRATES). Psychol Addict Behav 10:81–89.

Moos RH, Moos BS (2006) Rates and predictors of relapse after natural and treated remission from alcohol use disorders. Addiction 101:212–222.

National Institute on Alcohol Abuse and Alcoholism, US Department of Health and Human Services, National Institute of Health (2007) Helping Patients Who Drink Too Much: A Clinician's Guide (updated 2005 guide).

Orford J, Hawker A (1974) An investigation of an alcoholism rehabilitation halfway house: II the complex question of client motivation. Br J Addict Alcohol Other Drugs 69:315–323.

Orford J, Hodgson R, Copello A, John B, Smith M, Black R, Fryer K, Handforth L, Alwyn T, Kerr C, Thistlethwaite G, Slegg G (2006) The clients' perspective on change during treatment for an alcohol problem: qualitative analysis of follow-up interviews in the UK Alcohol Treatment Trial. Addiction 101:60–68.

Orford J, Hodgson R, Copello A, Wilton S, Slegg G (2008) To what factors do clients attribute change? Content analysis of follow-up interviews with clients of the UK Alcohol Treatment Trial. J Subst Abuse Treat 36:49–58.

Persson J, Magnusson P (1989) Early intervention in patients with excessive consumption of alcohol: a controlled study. Alcohol 6:403–408.

Pomerleau CS, Carton SM, Lutzke ML, Flessland KA, Pomerleau OF (1994) Reliability of the Fagerstrom Tolerance Questionnaire and the Fagerstrom Test for Nicotine Dependence. Addict Behav 19:33–39.

Rehm J, Room R, Graham K, Monteiro M, Gmel G, Sempos CT (2003) The relationship of average volume of alcohol consumption and patterns of drinking to burden of disease: an overview. Addiction 98:1209–1228.

Robins LN, Wing J, Wittchen HU, Helzer JE, Babor TF, Burke J, Farmer A, Jablenski A, Pickens R, Regier DA, Sartorius N, Towle L (1988) The Composite International Diagnostic Interview. An epidemiologic Instrument suitable for use in conjunction with different diagnostic systems and in different cultures. Arch Gen Psychiatry 45:1069–1077.

Room R, Babor T, Rehm J (2005) Alcohol and public health. Lancet 365:519–530.

Saitz R (2005) Clinical practice. Unhealthy alcohol use. N Engl J Med 352:596–607.

Saitz R, Freedner N, Palfai TP, Horton NJ, Samet JH (2006) The severity of unhealthy alcohol use in hospitalized medical patients. The spectrum is narrow. J Gen Intern Med 21:381–385.

Saitz R, Palfai TP, Cheng DM, Horton NJ, Dukes K, Kraemer KL, Roberts MS, Guerriero RT, Samet JH (2009) Some medical inpatients with unhealthy alcohol use may benefit from brief intervention. J Stud Alcohol Drugs 70:426–435.

Saitz R, Palfai TP, Cheng DM, Horton NJ, Freedner N, Dukes K, Kraemer KL, Roberts MS, Guerriero RT, Samet JH (2007) Brief intervention for medical inpatients with unhealthy alcohol use: a randomized, controlled trial. Ann Intern Med 146:167–176.

Samet JH, O'Connor PG (1998) Alcohol abusers in primary care: readiness to change behavior. Am J Med 105:302–306.

Smith KL, Horton NJ, Saitz R, Samet JH (2006) The use of the mini-mental state examination in recruitment for substance abuse research studies. Drug Alcohol Depend 82:231–237.

Smothers BA, Yahr HT, Sinclair MD (2003) Prevalence of current DSM-IV alcohol use disorders in short-stay, general hospital admissions, United States, 1994. Arch Intern Med 163:713–719.

Sobell LC, Sobell MB, Leo GI, Cancilla A (1988) Reliability of a Timeline Method: assessing normal drinkers' reports of recent drinking and a comparative evaluation across several populations. Br J Addict 83:393–402.

Solberg LI, Maciosek MV, Edwards NM (2008) Primary care intervention to reduce alcohol misuse ranking its health impact and cost effectiveness. Am J Prev Med 34:143–152.

StataCorp. (2007) Stata Statistical Software: Release Special Edition 10.1. Stata Corporation, College Station, TX.

Stein MD (1999) Medical consequences of substance abuse. Psychiatr Clin North Am 22:351–370.

Stewart SH, Connors GJ (2007) Perceived health status, alcohol-related problems, and readiness to change among medically hospitalized, alcohol-dependent patients. J Hosp Med 2:372–377.

Tsui JI, Saitz R, Cheng DM, Nunes D, Libman H, Alperen JK, Samet JH (2007) Awareness of hepatitis C diagnosis is associated with less alcohol use among persons co-infected with HIV. J Gen Intern Med 22:822–825.

Vielva I, Iraurgi I (2001) Cognitive and behavioural factors as predictors of abstinence following treatment for alcohol dependence. Addiction 96:297–303.

Walton MA, Blow FC, Bingham CR, Chermack ST (2003) Individual and social/environmental predictors of alcohol and drug use 2 years following substance abuse treatment. Addict Behav 28:627–642.

Walton MA, Goldstein AL, Chermack ST, McCammon RJ, Cunningham RM, Barry KL, Blow FC (2008) Brief alcohol intervention in the emergency department: moderators of effectiveness. J Stud Alcohol Drugs 69:550–560.

Ware J Jr, Kosinski M, Keller SD (1996) A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care 34:220–233.

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.

Willenbring ML, Olson DH (1999) A randomized trial of integrated outpatient treatment for medically ill alcoholic men. Arch Intern Med 159:1946–1952.

Williams EC, Horton NJ, Samet JH, Saitz R (2007) Do brief measures of readiness to change predict alcohol consumption and consequences in primary care patients with unhealthy alcohol use? Alcohol Clin Exp Res 31:428–435.

Williams EC, Kivlahan DR, Saitz R, Merrill JO, Achtmeyer CE, McCormick KA, Bradley KA (2006) Readiness to change in primary care patients who screened positive for alcohol misuse. Ann Fam Med 4:213–220.

Williams EC, Peytremann Bridevaux I, Fan VS, Bryson CL, Blough DK, Bradley KA (2010) The association between alcohol screening scores and self-reported health status in male veterans. J Addict Med 4:27–37.

Journal of Thrombosis and Haemostasis, **8**: 744–749 DOI: 10.1111/j.1538-7836.2010.03756.x

ORIGINAL ARTICLE

Twelve-month outcomes and predictors of very stable INR control in prevalent warfarin users

D. M. WITT,*†‡ T. DELATE,†‡ N. P. CLARK,*† C. MARTELL,† T. TRAN,† M. A. CROWTHER,§ D. A. GARCIA,¶ W. AGENO** and E. M. HYLEK†† ON BEHALF OF THE WARPED CONSORTIUM[1]

**Kaiser Permanente Colorado Clinical Pharmacy Anticoagulation Service, Lafayette, CO; †University of Colorado School of Pharmacy, Denver, CO; ‡Kaiser Permanente Colorado Clinical Pharmacy Research Team, Aurora, CO, USA; §McMaster University, Hamilton, Canada; ¶University of New Mexico School of Medicine, Albuquerque, NM, USA; **University of Insubria, Varese, Italy; and ††Boston University School of Medicine, Boston, MA, USA*

To cite this article: Witt DM, Delate T, Clark NP, Martell C, Tran T, Crowther MA, Garcia DA, Ageno W, Hylek EM, on behalf of the WARPED Consortium. Twelve-month outcomes and predictors of very stable INR control in prevalent warfarin users. *J Thromb Haemost* 2010; **8**: 744–9.

Summary. *Background:* For patients on warfarin therapy an international normalized ratio (INR) recall interval not exceeding 4 weeks has traditionally been recommended. For patients whose INR values are nearly always therapeutic, less frequent INR monitoring may be feasible. *Objective:* To identify patients with stable INRs (INR values exclusively within the INR range) and comparator patients (at least one INR outside the INR range), compare occurrences of thromboembolism, bleeding and death between groups, and identify independent predictors of stable INR control. *Methods:* The study was a retrospective, longitudinal cohort study using data extracted from electronic databases. Patient characteristics and risk factors were entered into multivariate logistic regression models to identify variables that independently predict stable INR status. *Results:* There were 533 stable and 2555 comparator patients. Bleeding and thromboembolic complications were significantly lower in stable vs. comparator patients (2.1% vs. 4.1% and 0.2% vs. 1.3%, respectively; $P < 0.05$). Independent predictors of stable INR control were age >70 years, male gender and the absence of heart failure. Stable patients were significantly less likely to have target INR ≥3.0 or chronic diseases. *Conclusion:* A group of patients with exclusively therapeutic INR values over 12 months is identifiable. In general, these patients are older, have a target INR <3.0, and do not have heart failure and/or other chronic diseases. Our findings suggest that many patients whose INR values remain within the therapeutic range over time could be safely treated with INR recall intervals >4 weeks.

Keywords: clinical outcomes, international normalized ratio, monitoring, warfarin.

Correspondence: Daniel M. Witt, Kaiser Permanente, Clinical Pharmacy Services, 16601 East Centretech Parkway, Aurora, CO 80011, USA.
Tel.: +1 303 739 3571; fax: +1 303 739 3574.
E-mail: dan.m.witt@kp.org

[1]Warfarin Associated Research Projects and other EnDeavors.

Received 11 September 2009, accepted 29 December 2009

Introduction

Warfarin is effective for the prevention and treatment of arterial and venous thromboembolic disorders. Intra- and interpatient variability in dose response, susceptibility to drug–drug and drug–food interactions and a narrow therapeutic index necessitate periodic monitoring of physiologic response to warfarin using the international normalized ratio (INR) [1]. In addition, to provide the best combination of thrombosis reduction and bleeding avoidance, target INR ranges are varied by therapeutic indication (e.g. an INR target range of 2.0–3.0 for atrial fibrillation or venous thrombosis and 2.5–3.5 for patients with mechanical heart valves) [1]. While the INR is used to monitor the impact of warfarin therapy, few studies have addressed the optimization of INR measurement frequency or INR recall interval. Current guidelines suggest an INR recall interval not exceeding 4 weeks between measurements [1,2]. However, this suggestion is not evidence-based, having evolved instead from regional differences in routine clinical practise and expert opinion [3].

Frequent INR testing, perhaps as often as weekly, has been suggested for patients who self-monitor warfarin using point-of-care technology [1,4,5]. High frequency INR testing raises the likelihood of measuring slightly out-of-range INR values (which often leads to unnecessary warfarin dose changes) [6], increases the costs associated with warfarin therapy, and is probably unnecessary in those patients who demonstrate long-term INR stability (i.e. minimal INR deviation and longitudinal warfarin dose stability). Less frequent INR monitoring should be possible for patients with a stable warfarin dose, as

suggested by routine clinical practise in the United Kingdom where INR recall intervals of up to 90 days are employed in such patients [7]. Recent evidence suggests that longer INR recall intervals may also be associated with improved INR control [8,9], which has in turn been associated with reduced risk for anticoagulation therapy-related adverse events [10,11]. Moreover, anticoagulated patients with 6-month stable INR control have been identified [12]. Patients with such stable INR control experienced significantly fewer anticoagulation therapy-related complications compared with anticoagulated patients who did not have 6 months of stable INR control.

The objective of this investigation was to expand on previous research by identifying anticoagulated patients with very stable (i.e. all INR values in the therapeutic range during a 12-month time interval) INR control. In addition to identifying patient characteristics associated with very stable INR values, we compared the rates of anticoagulation therapy-related adverse events with those from a group of comparator patients whose INR control was less stable.

Patients and methods

Study design and setting

The study was a retrospective, longitudinal cohort study conducted at Kaiser Permanente Colorado (KPCO), an integrated health care delivery system that provides services to over 450 000 members in the Denver–Boulder metropolitan area. Anticoagulation services at KPCO are provided by a centralized Clinical Pharmacy Anticoagulation Service (CPAS) [11]. Working collaboratively with the referring physician and using standardized dosing algorithms [13], CPAS clinical pharmacists initiate, adjust and refill anticoagulant medications and order relevant laboratory tests. Dosing algorithms utilized during this study at KPCO specified a maximum INR recall interval of 6 weeks. Integrated, electronic medical, pharmacy and laboratory records systems and the CPAS database (Dawn-AC®; 4S Systems, Ltd., Milnthorpe, UK) were utilized to identify patients, treatments and outcomes for this study. Approval to conduct this study was obtained from the KPCO Institutional Review Board.

Patients

Patients with a duration of warfarin therapy in excess of 90 days during the study timeframe (January 2000 through to December 2005), an age of greater than 18 years and warfarin therapy continuing throughout at least one 12-month observation period were included in the study. Stable patients were defined as having all INR values within each patient's strictly defined therapeutic INR range for the first identifiable continuous 12-month period (i.e. 100% INR control) during the study timeframe. Comparator patients were those who did not have any continuous 12-month period where all INR values were within the therapeutic range. In order to ensure a minimal standard for compliance with ongoing INR monitoring, both stable and comparator patients had to have at least one INR determination every 8 weeks during the 12-month observation period. For example, a patient who had an INR target range of 2.0–3.0, at least one INR measurement every 2 months for a minimum of six measurements, and all INR values recorded during the 12-month observation period between 2.0 and 3.0, would be assigned to the stable group.

Data collection

Variables collected for analysis included the primary warfarin indication, age at start of the 12-month observation period, gender, INR target, duration of warfarin therapy, and INR values. Patient-specific factors that could influence the risk for anticoagulant-related complications also were recorded, including a history of diabetes, hypertension, heart failure, venous thrombosis, hemorrhage, stroke, cancer, and/or use of estrogen therapy. Risk factors were considered present when a coded assessment for a given factor during a KPCO healthcare visit was identified in the 180 days prior to the start of the observation period. Estrogen therapy was defined as a prescription for a systemic estrogen-containing product sold within 90 days prior to the start of the observation period. A validated measure of patient acuity, the chronic disease score (CDS), was calculated for each patient by using ambulatory prescription drug data from the observation period [14]. Chronic disease scores can range from 0 to 35, with increasing scores indicating an increasing burden of chronic diseases under treatment. Use of the CDS allows for the accounting of each patient's risk of mortality and future health care utilization [14,15].

The first occurrence of anticoagulant-related complications (thromboembolism, bleeding and death) requiring admission to the hospital or emergency department were sought using ICD-9 discharge diagnostic codes (available upon request) within KPCO electronic administrative databases. All events were subsequently confirmed through independent review of the patient's electronic medical record by two investigators. Events were scored using a modified Naranjo Scale to quantify the relationship of the adverse event with warfarin therapy [16]. A third reviewer was employed to resolve disagreements.

Thromboembolic complications were defined as any deep vein thrombosis, pulmonary embolism, cerebral vascular accident, transient ischemic attack, systemic embolism, or heart valve thrombosis. Bleeding complications included episodes such as intracranial bleeding, gastrointestinal hemorrhage, hematoma, hemoptysis, epistaxis and hematuria. All bleeding episodes resulting in admission to the emergency department or hospital were included regardless of severity. Fatal events were assessed for direct relationship to bleeding or thromboembolism using the medical record and/or a death certificate.

Statistical analysis

Data analyses were performed utilizing SAS 9.1.3 (SAS, Cary, NC, USA) statistical software. Patient characteristics were

reported as means and standard deviations for interval-level variables (e.g. age, warfarin dose, length of warfarin therapy, etc.) and percentages for categorical variables (e.g. gender, target INR, occurrence of anticoagulation therapy-related complications, etc.). Group associations between categorical variables were assessed using the chi-square test and continuous variables were compared between groups using the independent samples *t*-test or Wilcoxon rank sum test (depending on the distribution of the data). Patient characteristics and risk factors with an overall incidence of ≥1% were entered into a multivariate logistic regression model to identify variables that independently predicted INR stability. The alpha was set at 0.05.

Results

Records from 4701 patients were screened; of these, 3088 patients had a period where at least one INR was measured every 8 weeks for 12 months. The stable group comprised 533 patients with INR values within the desired reference interval on all determinations and the comparator group comprised 2555 patients with at least one INR outside the desired reference interval.

Baseline characteristics of stable and comparator patients are presented in Table 1. Stable patients were older and more likely to have had a target INR < 3.0, heart failure or prior venous thrombosis, and been receiving warfarin for atrial fibrillation, compared with non-stable comparator patients (all $P < 0.05$). Stable patients were less likely to have been receiving warfarin for heart valve replacement, taking concurrent estrogen therapy, and their mean CDS was lower ($P < 0.05$). Differences in the duration of warfarin therapy prior to inclusion in the study were not statistically significant between groups ($P = 0.552$). The mean percentage of INR values in the therapeutic range for the comparator group was 42.1% [standard deviation (SD) = 5.7]. The stable patients had a lower mean number of INRs measured during the observation period, 12.6 (SD = 1.9) INRs per patient compared with 21.8 (SD = 8.0) for comparators ($P < 0.001$).

Compared with stable patients, the rate of overall mortality was higher in the comparator group ($P < 0.01$) (Table 2). However, the anticoagulation therapy-related mortality rates were not statistically significantly different ($P = 0.518$). The rates of anticoagulation-related thromboembolic ($P = 0.022$), bleeding ($P = 0.026$) and combined bleeding or thromboembolic ($P = 0.003$) complications were higher in the comparator group. Additionally, patients in the comparator group were more likely to require co-administration of heparin or low-molecular-weight heparin ($P < 0.001$).

Independent predictors of INR stability were age > 70 years [odds ratio (OR) = 1.93, 95% confidence interval (CI) 1.56–2.38], male sex (OR = 1.44, 95% CI 1.16–1.78), target INR of 2.0 (OR = 2.80, 95% CI 1.83–4.28), and the absence of co-morbid heart failure (OR = 2.08, 95% CI 1.36–3.17) (Table 3). Conversely, an increasing burden of chronic illness (i.e. higher CDS) (OR = 0.92, 95% CI 0.88–0.95) and a target INR ≥3.0 (OR = 0.28, 95% CI 0.17–0.47) were independent predictors of INR instability.

Table 1 Baseline characteristics

Characteristic	Stable group (n = 533)	Comparator group (n = 2555)	P-value
Mean age* (SD)	72.9 (9.5)	67.9 (13.5)	< 0.001
Age > 70 years (%)	65.5	48.7	< 0.001
Male (%)	58.7	50.0	< 0.001
INR target (%)			
2.0	7.5	2.5	< 0.001
2.5	86.1	77.2	< 0.001
≥3.0	6.4	20.3	< 0.001
Primary indication for anticoagulation therapy (%)			
Atrial fibrillation	51.0	41.4	< 0.001
Venous thromboembolism	23.3	26.0	0.196
Heart valve replacement	6.4	14.9	< 0.001
Other	19.3	17.8	0.410
risk factors (%)			
Diabetes mellitus†	2.4	4.2	0.061
Hypertension†	22.3	23.1	0.688
Heart failure†	5.1	10.6	< 0.001
Prior venous thrombosis†	2.4	4.3	0.050
Prior hemmorhage†	0.8	1.7	0.122
Prior stroke†	0.0	0.1	1.000
Cancer†	0.2	0.7	0.229
Estrogen therapy‡	8.4	11.5	0.040
Mean chronic disease score (SD)	6.2 (2.1)	6.7 (2.7)	< 0.001
Median duration of warfarin therapy (days)§ (IQR)	1220 (615, 1940)	790 (726, 1782)	0.552

INR, international normalized ratio; IQR, interquartile range; SD, standard deviation.*As of date of index INR measurement. † During the 180 days prior to the index INR. ‡During the 90 days prior to the index INR. §From initiation of warfarin therapy.

Discussion

In this large retrospective cohort study, we identified 533 patients with very stable long-term INR control. We found that patients with stable INR control were significantly less likely to develop complications related to anticoagulation therapy. We identified that age > 70 years, male gender, target INR of 2.0 and the absence of co-morbid heart failure independently predicted INR stability. Patients with a target INR ≥3.0 and those with a greater burden of chronic diseases were less likely to have long-term INR stability. On average the proportion of comparator patients' INRs in the therapeutic range was 42.1%. The time in therapeutic range for all patients managed by the KPCO CPAS is typically about 65% [11].

The results presented here are similar to those we observed in a study that evaluated patients with stable INR control over a 6-month observation period [12]. As would be expected based on a definition of stability that required exclusively therapeutic INRs, fewer patients were able to maintain 100% INR control

Table 2 Unadjusted outcomes during 365-day follow-up period

Characteristic	Stable group (n = 533)	Comparator group (n = 2555)	P-value
Received heparin* (%)	1.1	7.1	< 0.001
Deceased (n, %)	2, 0.4	51, 2.0	0.005†
AC-related death (n, %)	0, 0.0	2, 0.1	0.518†
AC-related thrombosis (n, %)	1, 0.2	34, 1.3	0.022†
Arterial thromboembolism	0, 0	1, 0.04	
Deep vein thrombosis	0, 0	4, 0.2	
Pulmonary embolism	0, 0	6, 0.2	
Stroke	1, 0.2	14, 0.5	
Thrombophlebitis	0, 0	1, 0.04	
Other	0, 0	8, 0.3	
AC-related bleeding (n, %)	11, 2.1	104, 4.1	0.026
Epistaxis	2, 0.4	24, 0.9	
Gastrointestinal	5, 0.9	44, 1.7	
Hemarthrosis	0, 0	3, 0.1	
Hematoma	0, 0	6, 0.2	
Hematuria	1, 0.2	10, 0.4	
Intracranial	1, 0.2	8, 0.3	
Other	2, 0.4	9, 0.4	
AC-related bleeding or thrombosis (n, %)	12, 2.3	136, 5.3	0.003

AC, anticoagulation. *Heparin or low-molecular-weight heparin. †Fisher's exact test.

Table 3 Predictors of stable INR control status (c-statistic = 0.69)

Predictor	Adjusted odds ratio	95% CI
Age		
> 70 years	1.93	1.56–2.38
≤70 years	–	–
Sex		
Female	–	–
Male	1.44	1.16–1.78
INR target		
2.0	2.80	1.83–4.28
2.5	–	–
≥3.0	0.28	0.17–0.47
Primary indication for anticoagulation therapy		
Atrial fibrillation	–	–
Venous thromboembolism	0.81	0.63–1.04
Heart valve disorder	1.13	0.65–1.98
Other	1.01	0.77–1.31
Risk factors		
Diabetes mellitus		
Yes	–	–
No	1.69	0.93–3.08
Hypertension		
Yes	–	–
No	0.98	0.77–1.24
Heart failure		
Yes	–	–
No	2.08	1.36–3.17
Prior venous thrombosis		
Yes	–	–
No	1.53	0.84–2.79
Prior hemorrhage		
Yes	–	–
No	2.02	0.70–5.79
Estrogen therapy		
Yes	–	–
No	1.04	0.72–1.49
Chronic disease score*	0.921	0.88–0.95

*With all things being equal between two patients, for every 1 point increase in chronic disease score there is an 8% reduced likelihood of being stable.

over 12 months compared with 6 months of observation (533 vs. 2504, respectively). However, both studies found that age > 70, absence of heart failure, target INR < 3.0 and a lesser burden of chronic illness independently predicted stable INR control, increasing confidence in the observed association of these variables with INR stability. The two analyses differ in that male gender was identified as an independent predictor of stability in the current study, while estrogen therapy was not.

Based on the mean number of INRs measured in stable patients during the 12-month observation period (12.6), the observed frequency of INR monitoring was consistent with the current North American standard 4-week recall interval in stable patients. However, as study patients were required to have at least one INR measured every 8 weeks, it is reasonable to conclude in retrospect that stable group patients could have been safely monitored with 8-week INR recall intervals; so doing would have reduced the count of INR measurements by approximately half without compromising therapeutic outcomes. Less frequent INR monitoring would probably be well received by patients on chronic warfarin therapy and would reduce the utilization of healthcare resources. Changes in health status, dietary vitamin K intake or co-administered medications would require more frequent INR monitoring [1]. The utility of 8-week recall intervals should be confirmed in prospective, controlled trials. The observation that comparator patients were monitored significantly more frequently (mean INRs per patient 21.8) than stable patients probably reflects the standard CPAS procedure of scheduling 1–2-week INR recall intervals for out-of-range INR values. While it is possible that more frequent interaction with the healthcare system in comparator patients increased the chances of identifying bleeding or thromboembolic complications, we believe this is unlikely. Our definition of complications required admission to the hospital or emergency department (see Table 2), and most encounters for INR testing in our system involve only a visit to the laboratory for phlebotomy and do not include formal evaluation by a healthcare provider.

The finding that age > 70 predicted long-term INR stability argues against innate INR variability associated with advancing age. Although not specifically assessed, we speculate that better adherence to warfarin treatment regimens among older patients may have been a factor. We do not fully understand the association between a lower target INR (2.0) and INR stability. However, this observation could merely be a consequence of the non-linear association between INR and clotting factor activity [17]. Thus increasing INR target intervals of equal width such as those observed in our study (e.g. 1.5–2.5, 2.0–3.0, 2.5–3.5) may in effect have narrower therapeutic intervals. This is consistent with our observation that INR

targets of ≥3.0 independently predicted INR instability. For patients with atrial fibrillation, INR values < 2.0 have been associated with an increased risk for thromboembolic complications and more serious sequelae from these complications [18,19]. In light of this, and the absence of high-quality evidence showing that a target INR = 2.0 significantly lowers the risk of major hemorrhage, we recommend against targeting a lower INR range as a way to extend INR recall intervals.

Our results are likely to be valid. The dataset used to complete this study was robust, includes real-world patients with a variety of indications for warfarin and therapeutic INR targets, and has been used previously in health records and data extraction research [11,12,20] The large number of patients included in our analysis increases the generalizability of our results and reduces the likelihood that unmeasured bias may have influenced them. Clinical events were comprehensively collected and described and INR measurements were performed by a single laboratory and systematically captured in an integrated electronic medical record. All clinical events were independently assessed for causality using a validated scoring system by two expert reviewers. The long-term stable cohort was carefully established using a strict definition for stability. Most patients observed in our study had been on warfarin therapy for several years. Potential adherence and survivor biases were minimized by the fact that both groups observed in our study could appropriately be termed 'prevalent' warfarin users.

This study does have important limitations. Not all variables likely to enter into clinical decision making or known to affect INR control were collected due to the retrospective, observational design. The retrospective nature of the study also precludes definitive establishment of cause and effect relationships between study variables and outcomes. It is possible that some clinical events were missed. However, given that KPCO patients are either seen within an affiliated hospital or the costs of care are billed to KPCO when care is provided at non-affiliated hospitals, it is likely that the vast majority of clinically important events were captured. Any failure to capture such events would have been random across groups. Our study was conducted within an integrated healthcare delivery system with a specialized anticoagulation service using standardized warfarin dosing protocols and, thus, the observed results may not directly translate to other healthcare settings.

In conclusion, our work proves that a subgroup of anticoagulated patients with therapeutically stable INR values over 12 months can be identified. In general, these patients are older, with a target INR < 3.0 and lower chronic illness burden. Additionally, patients with highly stable INR values appear to experience significantly fewer anticoagulation therapy-related complications. Our findings support the suggestion that INR recall intervals should be individually tailored based on demonstrated INR control rather than being fixed at some arbitrary minimum frequency, such as 4 weeks [8]. We acknowledge that, before widespread change in the frequency of INR monitoring can be recommended, our findings will need confirmation in future prospective evaluations.

Addendum

D.M. Witt, T. Delate and N.P. Clark designed the research and extracted information from medical records; T. Delate performed the statistical analysis; D.M. Witt, T. Delate, N.P. Clark, M.A. Crowther, D.A. Garcia, W. Ageno and E.M. Hylek interpreted the analysis and revised the manuscript; D.M. Witt wrote the initial draft of the manuscript; T. Tran and C. Martell extracted information from medical records and reviewed the manuscript.

Acknowledgements

This work was funded entirely by Kaiser Permanente Colorado.

Disclosure of Conflict of Interests

E.M. Hylek reports serving as an advisor to Bayer, Boehringer Ingelheim, Bristol-Myers Squibb and Sanofi-Aventis, and participating in clinical symposia sponsored by Bayer, Boehringer Ingelheim, and Bristol-Myers Squibb. D.A.G. reports serving as a consultant to Roche Diagnostics. The remaining authors state that they have no conflict of interest.

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**: 160–98.

2 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, *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.

3 Fitzmaurice DA. Oral anticoagulation therapy should be managed in the community with treatment aimed at standard targets and increased recal intervals. *J Thromb Haemost* 2008; **6**: 1645–6.

4 Heneghan C, Alonso-Coello P, Garcia-Alamino M, Perera R, Meats E, Glasziou P. Self-monitoring of oral anticoagulation: a systematic review and meta-analysis. *Lancet* 2006; **367**: 404–11.

5 Horstkotte D, Piper C, Wiemer M. Optimal frequency of patient monitoring and intensity of oral anticoagulation therapy in valvular heart disease. *J Thromb Thrombolysis* 1998; **5** (Suppl. 1): 19–24.

6 Rose AJ, Ozonoff A, Berlowitz DR, Henault LE, Hylek EM. Warfarin dose management affects INR control. *J Thromb Haemost* 2009; **7**: 94–101.

7 Guidelines on oral anticoagulation: third edition. *Br J Haematol* 1998; **101**: 374–87.

8 Rose AJ, Ozonoff A, Henault LE, Hylek EM. Warfarin for atrial fibrillation in community-based practise. *J Thromb Haemost* 2008; **6**: 1647–54.

9 Snyder CM, Helms BE, Hall DL. Evaluation of INR monitoring frequency and time in therapeutic range. *J Pharm Technol* 2008; **24**: 255–60.
10 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**: 283–92.
11 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**: 1515–22.
12 Witt DM, Delate T, Clark NP, Martell C, Tran T, Crowther MA, Garcia DA, Ageno W, Hylek EM. Outcomes and predictors of very stable INR control during chronic anticoagulation therapy. *Blood* 2009; **114**: 952–6.
13 Poller L, Shiach CR, MacCallum PK, Johansen AM, Münster AM, Magalhães A, Jespersen J. Multicentre randomised study of computerised anticoagulant dosage. European Concerted Action on Anticoagulation. *Lancet* 1998; **352**: 1505–9.
14 Clark DO, Von Korff M, Saunders K, Baluch WM, Simon GE. A chronic disease score with empirically derived weights. *Med Care* 1995; **33**: 783–95.
15 Von Korff M, Wagner EH, Saunders K. A chronic disease score from automated pharmacy data. *J Clin Epidemiol* 1992; **45**: 197–203.
16 Naranjo CA, Busto U, Sellers EM, Sandor P, Ruiz I, Roberts EA, Janecek E, Domecq C, Greenblatt DJ. A method for estimating the probability of adverse drug reactions. *Clin Pharmacol Ther* 1981; **30**: 239–45.
17 Yuan S, Ferrell C, Chandler WL. Comparing the prothrombin time INR versus the APTT to evaluate the coagulopathy of acute trauma. *Thromb Res* 2007; **120**: 29–37.
18 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**: 540–6.
19 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.
20 Clark NP, Witt DM, Delate T, Trapp M, Garcia D, Ageno W, Hylek EM, Crowther MA. Thromboembolic consequences of subtherapeutic anticoagulation in patients stabilized on warfarin therapy: the low INR study. *Pharmacotherapy* 2008; **28**: 960–7.

Quality of US Primary Care Delivered by Resident and Staff Physicians

Leah Zallman, MD[1], Jun Ma, MD, PhD[2], Lan Xiao, PhD[2], and Karen E. Lasser, MD, MPH[3]

[1] Department of Medicine, Cambridge Health Alliance and Harvard Medical School, Cambridge, MA, USA; [2]Palo Alto Medical Foundation Research Institute, Palo Alto, CA, USA; [3]Section of General Internal Medicine, Boston Medical Center, Boston, MA, USA.

BACKGROUND: Few population-based data are available on the quality of outpatient care provided by resident physicians in the US.

OBJECTIVE: To assess the quality of outpatient care delivered by resident and staff physicians.

DESIGN: Cross-sectional analysis. We used chi-square tests to compare resident and staff physician performance on 19 quality indicators. Using multivariable logistic regression, we controlled for sex, age, race/ethnicity, insurance, and metropolitan status.

PARTICIPANTS: 33,900 hospital-based outpatient visits from the 1997-2004 National Hospital Ambulatory Medical Care Survey (NHAMCS).

MEASUREMENTS: Resident and staff physician performance on 19 quality indicators.

RESULTS: Resident physicians were more likely to care for younger, non-white, female, urban, and Medicaid-insured patients. In both adjusted and unadjusted analyses, residents outperformed staff on four of 19 measures including angiotensin converting enzyme inhibitor use for congestive heart failure (57.0% vs. 27.6%; p=<0.001), diuretic use for hypertension (57.8% vs. 44.0%; p=<0.001), statin use for hyperlipidemia (56.3% vs. 40.4%; p=0.001), and routine blood pressure screening (85.3% vs. 79.6%; p=0.02). Residents and staff performed at similar levels for counseling (range 15.7 to 32.0%). Residents and staff performed similarly well on measures capturing inappropriate prescribing or overuse of diagnostic testing (range 48.6 to 100%). Residents and staff performed similarly on measures of appropriate prescribing (range from 30.9% to 69.2%).

CONCLUSIONS: Primary care provided by resident physicians is of similar or higher quality than that provided by staff physicians. Significant opportunity remains to improve quality of outpatient care provided by all physicians. Residency training programs should devote attention to improving outpatient quality of care delivered by residents.

KEY WORDS: resident; quality of care; primary care.

Abbreviations

ACE	angiotensin converting enzyme
BB	beta blocker
CAD	coronary artery disease
CBC	complete blood count
CHF	chronic heart failure
ECG	electrocardiography
HTN	hypertension
IC	inhaled corticosteroid
TMP SMX	trimethoprim sulfamethoxazole
URTI	upper respiratory tract infection
UTI	urinary tract infection

Received March 4, 2010
Revised June 18, 2010
Accepted July 6, 2010
Published online July 20, 2010

J Gen Intern Med 25(11):1193–7
DOI: 10.1007/s11606-010-1456-0

BACKGROUND

Despite ongoing efforts to improve the quality of primary care in the United States, deficiencies persist.[1] In recognition of the importance of residency training in ambulatory medicine, the Accreditation Council for Graduate Medical Education (ACGME) increased the required time residents spend in primary care continuity clinics starting in July 2009 [2] and has supported involvement of residents in quality improvement efforts. Targeting trainees is prudent, as practice patterns are largely determined during the residency training years. Quality of care delivered by internal medicine residents is particularly important with respect to addressing health care disparities, as residents provide a substantial amount of care to minority and uninsured patients.[3,4]

Studies have shown increased quality of inpatient care delivered in teaching hospitals relative to non-teaching hospitals.[5,6] However, few population-based data are available on the quality of outpatient care provided by resident physicians in the United States. Although prior studies have reported comparable or increased quality of outpatient care delivered by residents, these studies have been limited by their lack of a comparison group [7] or by their focus on specific diseases or geographical areas.[8,9] We analyzed data from the National Hospital Ambulatory Medical Care Survey (NHAMCS). Our objective was to compare the quality of outpatient care delivered by resident physicians and staff physicians in the United States, across a spectrum of previously defined and published quality measures.[10]

METHODS

Data Source

We analyzed data from the Outpatient Department component of the NHAMCS collected between 1997 and 2004. The NHAMCS is administered by the National Center for Health Statistics (NCHS). It uses a multistage stratified probability sampling procedure to obtain nationally representative estimates of outpatient visits to hospital-based clinics on an annual basis. Hospital-based physicians (with staff assistance) or hospital staff complete NCHS standard encounter forms, including data on visit diagnoses, over a four-week period. The survey contains items on patient demographics (with the patient's race and ethnicity determined by the physician), up to three reasons for the visit (in the patient's own words), visit diagnoses (primary diagnosis plus two 'other' diagnoses), diagnostic and screening services ordered, counseling and education provided, and up to six medications. Listed medications included prescription and nonprescription medications that the physician prescribed on the day of the visit or prior to the visit and that the physician expected the patient to continue taking. In addition, the physician indicates whether she is the patient's primary care physician and whether she is a staff or resident physician. We did not analyze NHAMCS data after 2004 because the later surveys did not identify resident physicians. Ninety percent of the selected hospitals participated, and item non-response rates were generally 5% or less.

Description of Quality Indicators

We examined a set of 20 outpatient quality indicators, which was previously described in detail for use in NHAMCS.[10] Briefly, the quality indicators were developed in accordance with the Institute of Medicine's criteria[11] of clinical importance, scientific soundness, and feasibility for indicator selection as well as criteria specific to the limitations of the data source. Limitations of the data source included visit-based data, unreliable data elements and subpopulations, and inconsistent inclusion of variables over years. Indicators were selected if they were meaningful when measured at individual patient visits and did not rely on data elements or subpopulations that are considered unreliable according to NCHS standards. The quality indicators fall into five categories: 1) medical management of chronic diseases (ten measures; e.g. beta blocker use for coronary artery disease); 2) appropriate antibiotic use (two measures; e.g. no antibiotic use for upper respiratory tract infection); 3) preventive counseling (three measures; e.g. exercise counseling in adults at moderate to high risk for coronary heart disease defined as having two or more risk factors including a history of smoking, men >45 years old or women >55 years old, hypercholesterolemia, hypertension, or obesity and excluding those with known CAD or diabetes mellitus); 4) screening tests (four measures; e.g. routine blood pressure screening), and 5) inappropriate prescribing in elderly patients, defined as age $\geq$ 65 (e.g. avoiding use of 33 inappropriate medications such as benzodiazepines). We defined performance on quality indicators as the percentage of eligible visits receiving recommended care (i.e., the higher the percentage the better the performance), based on practice guidelines or, in the absence of authoritative practice guidelines, consensus expert statements. We excluded visits from both the numerator and denominator of a measure if the patient had clinical contraindications to a recommended treatment. For example, we computed the measure 'angiotensin converting enzyme (ACE) inhibitor use for congestive heart failure (CHF)' as the number of visits by adults with a visit diagnosis of CHF who had a documented prescription of ACE inhibitors or angiotensin receptor blockers divided by the number of visits by adults diagnosed as having CHF. We excluded visits by adults with diagnoses of hyperkalemia or angioedema from both the numerator and the denominator. We could only capture disease conditions and exclusions if the diagnosis in question was listed on the day of the visit, and if the medication in question was listed on the day of the visit. Because this study used publicly available anonymous data, the Institutional Review Board of Boston University Medical Center deemed it exempt from review.

Statistical Analysis

Using SAS Enterprise Guide 4.2 (SAS Institute, Cary, NC), we performed analyses using the patient visit as the unit of analysis. We calculated standard errors for all results as recommended by the NCHS, which accounts for the sampling weights and the complex multi-stage sampling design of the NHAMCS.[12] According to the NCHS, estimates with greater than a 30% relative standard error (i.e. the standard error divided by the estimate expressed as a percentage of the estimate) or based on fewer than 30 sample cases may be unreliable. In accordance with the NCHS analytical guidelines, we combined data from 1997-2004 in order to generate reliable national estimates. We excluded one quality indicator (antithrombotic therapy for atrial fibrillation), because it had a sample size that was too small to be considered reliable, leaving us with 19 quality indicators.

The primary outcomes were mean performance rates on quality indicators analyzed according to physician status (resident vs. staff physician). We compared mean performance rates with chi-square analyses (PROC SURVEYFREQ), followed by multivariable logistic regression analyses (PROC SURVEYLOGISTIC) controlling for patient sex, age, race/ethnicity, medical insurance, and metropolitan status. We selected these variables because they have been shown to be associated with differences in quality of care.[13,14]

RESULTS

Patient and Visit Characteristics

Table 1 shows characteristics of patient visits made to resident and staff physicians. Resident physicians were more likely to care for younger, non-white, female, and urban patients, as well as patients with Medicaid. Resident and staff physicians cared for patients with similar numbers of comorbid medical conditions and from similar regions of the United States. The proportion of visits to resident and staff physicians for preventive care versus treatment was similar.

Table 1. Demographic Characteristics of All Patient Visits to U.S. Resident and Staff Primary Care Physicians as Percentages, 1997-2004

Variable	Resident Physician	Staff Physician	χ2 P value
	Visits	Visits	
	n=6322	n=27578	
	%	%	
Patient characteristics			
Sex			
Women	68.5	64.8	0.019
Age			
20-44	46.8	38.7	<0.001
45-64	37.4	35.2	
>=65	15.8	26.2	
Race			
White	52.2	67.8	<0.001
Black/African American	34.5	17.8	
Hispanic	11.4	11.2	
Other	1.9	3.2	
Comorbidity[a]			
Yes	34.7	31.6	NS
Mean number	0.49±0.06	0.42±0.02	NS
Visit characteristics			
Medical insurance			
Private	19.5	38.0	<0.001
Government-sponsored			
Medicaid/SCHIP	31.8	19.8	
Medicare	17.4	24.5	
Other[†]	31.2	17.7	
Visit type			
Preventive care	19.1	16.0	NS
Treatment visit	80.9	84.0	
Geographic region			
Northeast	29.2	27.6	NS
Midwest	33.6	29.0	
South	29.4	33.8	
West	7.8	9.6	
Metropolitan statistical area			
Yes	99.0	79.0	<0.001

[a]Comorbidity defined as coronary heart disease (*n*=967), diabetes (*n*=3554), hypertension (*n*=6469), hypercholesterolemia (*n*=1504), congestive heart failure (*n*=507) or stroke (*n*=207)
[†]"Other" denotes self-pay, no charge/charity, worker's compensation, other and unknown

Table 2. Quality of Outpatient Care Among All Patient Visits to US Resident and Staff Physicians, 1997-2004

	Resident Physician	Staff Physician	χ2 P value
	Visits	Visits	
	% (95% CI)	% (95% CI)	
Medical Management of Common Diseases (*N*=number of visits associated with issue)			
ACE inhibitor use for CHF (N=506)	57.0 (39.6,74.4)	27.6 (19.2,36.0)	<0.001[a]
Aspirin use for CAD (N=944)	48.4 (25.1,71.7)	35.8 (27.3,44.2)	NS
BB for CAD (N=903)	31.9 (20.4,43.4)	30.9 (23,38.9)	NS
Diuretic use for HTN (N=3020)	57.8 (51.5, 64.1)	44.0 (40.5,47.6)	<0.001[a]
IC use for asthma in adults (N=945)	38.2 (26.5, 49.8)	32.3 (24.7, 39.6)	NS
Statin use for hyperlipidemia (N=1269)	56.3 (48.3, 64.2)	40.4 (34.9, 45.8)	0.001[a]
Treatment of depression[†] (N=1783)	65.2 (52.6, 77.9)	69.2 (63.8, 74.6)	NS
No benzodiazepine use for depression (N=1783)	100 (100, 100)	100 (99.9,100)	NS
Hemoglobin A1C checked for diabetes[§] (N=922)	23.3 (10.2, 36.4)	22.0 (13.1-30.9)	NS
Preventive Counseling			
Smoking cessation (N=493)	20.7 (8.7,32.6)	25.6 (17.7,33.4)	NS
Diet/nutrition in high-risk adults[‡] (N=732)	32.0 (20.7,43.2)	28.3 (21.7,34.9)	NS
Exercise in high-risk adults[‡] (N=732)	16.3 (7.1,25.5)	15.7 (9.6,21.9)	NS
Screening Tests			
Blood pressure screen (N=8061)	85.3 (81.1,89.5)	79.6 (76.4,82.8)	0.024[a]
No routine ECG (N=1335)	95.3 (91.6,99.0)	91.5 (87.6,95.3)	NS
No routine urinalysis (N=3026)	48.6 (38.5,58.7)	53.7 (47.0,60.3)	NS
No routine CBC (N=551)	88.0 (74.5,100)	81.5 (74.0,89.1)	NS
Appropriate Antibiotic Use			
TMP-SMX or quinolone use for UTI (N=345)	62.2 (45.8,78.6)	60.3 (49.9,70.7)	NS
No antibiotic use for URTI (N=348)	81.5 (69.0,94.0)	78.0 (69.5, 86.7)	NS
Inappropriate Prescribing in Elderly Patients			
Avoiding potentially inappropriate prescribing in elderly patients (N=7372)	93.8 (91.4, 96.1)	93.2 (91.9,94.6)	NS

[a]denotes that significance remains after adjusting for sex, age, race, insurance, and urban location
[†]treatment of depression is defined as: prescribing antidepressants, psychotherapy or mental health counseling
[‡]high risk adults is defined as having two or more of the following risk factors for coronary heart disease: a history of smoking, men >45 years old or women >55 years old, hypercholesterolemia, hypertension, or obesity and excluding those with known CAD or diabetes mellitus
[§]Data only available for 2003 and 2004

Performance on Quality Indicators

Table 2 shows mean performance rates on quality indicators for resident and staff physicians. In both unadjusted and adjusted analyses, residents outperformed staff physicians on four measures: ACE inhibitor use for congestive heart failure (57.0% vs. 27.6%; adjusted p=<0.001), diuretic use for hypertension (57.8% vs. 44.0% adjusted p=<0.001), statin use for hyperlipidemia (56.3% vs. 40.4% adjusted p=0.001), and routine blood pressure screening (85.3% vs. 79.6%; adjusted p=0.024). Residents and staff performed at similar levels for counseling on smoking cessation, and for exercise and nutrition counseling among moderate to high-risk adults; performance rates ranged from 15.7 to 32.0%. Residents and staff performed similarly well on measures capturing inappropriate prescribing or overuse of diagnostic testing. These measures included avoiding inappropriate prescribing in elderly patients, benzodiazepine use for the treatment of depression, antibiotic use for upper respiratory tract infections, and ordering routine complete blood counts, urinalyses and electrocardiograms; performance rates varied from 48.6% to 100%. Residents

and staff performed similarly on measures of appropriate prescribing including: treatment of depression (prescribing antidepressants, psychotherapy or mental health counseling), trimethoprim-sulfamethoxazole or quinolone use for urinary tract infections, aspirin and beta blocker use for coronary artery disease, and inhaled corticosteroid use for asthma. Performance rates on these measures of appropriate prescribing were between 30.9% and 69.2%.

COMMENT

Compared with staff physicians, resident physicians provide primary care that is of similar or higher quality. Residents outperformed staff physicians on four of nineteen quality indicators including ACE inhibitor use for congestive heart failure, diuretic use for hypertension, statin use for hyperlipidemia and routine blood pressure screening. Residents and staff perform similarly on preventive counseling, measures of avoiding inappropriate prescribing or overuse of diagnostic testing, and on measures of appropriate prescribing.

Our results are congruent with previous studies that have demonstrated increased quality of care provided by resident physicians on specific measures or at single sites.[8,9] As in previous studies, we found that resident physicians are less likely than staff physicians to prescribe antibiotics for respiratory conditions where antibiotics are rarely indicated.[8] Both resident and staff physician performance revealed deficits in quality of care. This confirms previously identified gaps in compliance with evidence-based guidelines for blood pressure screening,[15] diet and exercise counseling,[16] statin use for moderate to high risk patients,[17] and aspirin use for primary and secondary prevention of cardiovascular disease.[1,18,19]

Resident physicians provide a disproportionate amount of care for vulnerable populations. In our study, residents were more likely than staff physicians to care for non-white and Medicaid patients, which is consistent with prior studies.[3,4] This finding suggests that improving the quality of resident outpatient care might improve the quality of care delivered to underserved populations and hence potentially reduce disparities in care.

Our analyses are limited because physicians were only able to list a limited number of diagnoses (up to three) and medications (up to six). Thus, some diagnoses and medications may not have been captured on the day of the NHAMCS study visit. The six-medication limit would cause us to underestimate quality for indicators that call for use of a specific drug, but overestimate quality when the measure suggests drug avoidance. Hence, we would expect measures capturing inappropriate prescribing (avoiding inappropriate prescribing in elderly patients, benzodiazepine use for the treatment of depression, and antibiotic use for upper respiratory tract infections) to be somewhat overestimated in our comparisons, while measures of appropriate prescribing (treatment of depression, trimethoprim-sulfamethoxazole or quinolone use for urinary tract infections, aspirin and beta blocker use for coronary artery disease, and inhaled corticosteroid use for asthma) would be underestimated. Similarly, lack of complete diagnostic information might cause us to underestimate valid contraindications to otherwise indicated drugs. This problem seems most likely for ACE inhibitor in CHF, beta-blocker for CAD, diuretic use for HTN, and statin use for hyperlipidemia. However, these problems would only affect our conclusions regarding the relative quality of care by the two groups of physicians if these limitations applied differentially to residents or staff physicians. We know of no a priori reason to believe this is so and in fact believe that patients cared for by resident physicians have a similar burden of illness as those cared for by staff physicians, as evidenced by our finding that patients in each group were as likely to have a comorbid condition and had a similar number of comorbid conditions. Thus, we doubt that the absence of data on additional diagnoses and medications has introduced bias.

Our analyses are also limited by underreporting of preventive counseling behaviors in physician-report-based studies such as the NHAMCS, in which counseling behaviors are underreported for two reasons: 1) physicians underreport counseling behavior as compared to procedural behaviors,[20] and 2) visit-based data are unable to capture counseling behaviors that did not occur at the index visit. We doubt that staff physicians are more likely than residents to underreport counseling behaviors; however, because staff physicians have most likely seen their patients over a longer period of time, counseling is more likely to have occurred at a prior visit. One of the quality measures, blood pressure screening, is often carried out by medical assistants and may therefore reflect a difference in practice performance rather than individual physician performance. However, as it is incumbent upon the physician to ensure that screening tests be performed, the difference in blood pressure screening may still reflect a difference in physician attention to this screening test. Our study data were limited to hospital-based clinics and do not include visits to community-based clinics. Thus our findings may not be generalizable to those settings. Finally, because our study period preceded the ACGME-mandated increased ambulatory requirements, we have not captured any improvements in quality of care that may have resulted from this change.

Why are there differences in the quality of outpatient care provided by resident and staff physicians? Resident physicians are in the midst of training, thereby increasing the likelihood that they have learned current evidence-based based guidelines. Residents are also closely precepted by faculty, who may be more likely than non-teaching staff physicians to practice according to current guidelines. Residents performed particularly well on measures that are reinforced by their inpatient training, such as ACE inhibitor use for CHF and statin use for hyperlipidemia. Finally, since we were unable to control for the type of institution (academic vs. non-academic), the higher quality of care delivered by residents may reflect differences in systems of care provided at academic versus non-academic institutions.

This study highlights the significant opportunity that remains to improve quality of outpatient care in the United States, including within residency training programs. While ongoing quality improvement projects aimed at staff physicians target the majority of care delivered in the US, quality improvement programs targeting residency training sites are an important area for future interventions. Despite the fact that the majority of care is delivered in ambulatory settings, residents spend the majority of their training years in inpatient settings. Residency training programs should devote attention not only to augmenting time spent in the ambulatory setting as mandated by the ACGME but also on improving outpatient quality of care delivered by residents. Residents might participate in quality improvement projects in order to learn more about quality measurement and process improvement. Such efforts could be

guided by the development of a standard curriculum. As more emphasis is being placed on performance-based compensation, this is an increasingly important area to which residency programs should devote attention. Quality improvement interventions that address specific barriers at the system, provider, and patient level are necessary in order to reduce the discrepancy between clinical practice and best evidence.[21,22] Research on factors contributing to deficiencies and disparities in quality of care will inform the design of tailored interventions.

Acknowledgements: *We thank Dr Stephanie Woolhandler for her editorial comments, and Meredith Manze for her assistance with manuscript preparation.*

Conflict of Interest: *There are no financial conflicts of interest for any of the authors.*

Corresponding Author: *Karen E. Lasser, MD, MPH, Section of General Internal Medicine, Boston Medical Center, 801 Massachusetts Avenue, Crosstown #2091, Boston, MA 02118, USA (e-mail: Karen.lasser@bmc.org).*

REFERENCES

1. **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(26):2635–45.
2. ACGME. http://www.acgme.org/acWebsite/downloads/RRC_progReq/140_internal_medicine_07012009_TCC.pdf. Accessed 7/6//10.
3. **Charlson ME, Karnik J, Wong M, McCulloch CE, Hollenberg JP.** Does experience matter? A comparison of the practice of attendings and residents. J Gen Intern Med. 2005;20(6):497–503.
4. **Serwint JR, Thoma KA, Dabrow SM, et al.** Comparing patients seen in pediatric resident continuity clinics and national ambulatory medical care survey practices: a study from the continuity research network. Pediatrics. 2006;118(3):e849–858.
5. **Kupersmith J.** Quality of care in teaching hospitals: a literature review. Acad Med. 2005;80(5):458–66.
6. **Ayanian JZ, Weissman JS.** Teaching hospitals and quality of care: a review of the literature. Milbank Q. 2002;80(3):569–93. v.
7. **Mladenovic J, Shea JA, Duffy FD, Lynn LA, Holmboe ES, Lipner RS.** Variation in internal medicine residency clinic practices: assessing practice environments and quality of care. J Gen Intern Med. 2008;23(7):914–20.
8. **Roumie CL, Halasa NB, Edwards KM, Zhu Y, Dittus RS, Griffin MR.** Differences in antibiotic prescribing among physicians, residents, and nonphysician clinicians. Am J Med. 2005;118(6):641–8.
9. **Tsui JI, Dodson K, Jacobson TA.** Cardiovascular disease prevention counseling in residency: resident and attending physician attitudes and practices. J Natl Med Assoc. 2004;96(8):1080–3. 1088-1091.
10. **Ma J, Stafford RS.** Quality of US outpatient care: temporal changes and racial/ethnic disparities. Arch Intern Med. 2005;165(12):1354–61.
11. **Hurtado MPSE, Corrigan JM.** Committee on the National Quality Report on Health Care Delivery, Board on Health Care Services. *Envisioning the National Health Care Quality Report*. Washington: National Academy Press; 2001.
12. CDC. http://www.cdc.gov/nchs/ahcd.htm Accessed 7/6/10.
13. **Asch SM, Kerr EA, Keesey J, et al.** Who is at greatest risk for receiving poor-quality health care? N Engl J Med. 2006;354(11):1147–56.
14. **Zaslavsky AM, Hochheimer JN, Schneider EC, et al.** Impact of sociodemographic case mix on the HEDIS measures of health plan quality. Med Care. 2000;38(10):981–92.
15. **Ma J, Stafford RS.** Screening, treatment, and control of hypertension in US private physician offices, 2003-2004. Hypertension. 2008;51(5):1275–81.
16. **Ma J, Xiao L, Stafford RS.** Adult obesity and office-based quality of care in the United States. Obesity. 2009;17(5):1077–85.
17. **Ma J, Sehgal NL, Ayanian JZ, Stafford RS.** National trends in statin use by coronary heart disease risk category. PLoS Med. 2005;2(5):e123.
18. **Stafford RS, Monti V, Ma J.** Underutilization of aspirin persists in US ambulatory care for the secondary and primary prevention of cardiovascular disease. PLoS Med. 2005;2(12):e353.
19. **Schuster MA, McGlynn EA, Brook RH.** How good is the quality of health care in the United States? Milbank Q. 1998;76(4):517–63. 509.
20. **Gilchrist VJ, Stange KC, Flocke SA, McCord G, Bourguet CC.** A comparison of the National Ambulatory Medical Care Survey (NAMCS) measurement approach with direct observation of outpatient visits. Med Care. 2004;42(3):276–80.
21. **Bergeson SC, Dean JD.** A systems approach to patient-centered care. Jama. 2006;296(23):2848–51.
22. **Wagner EH, Austin BT, Davis C, Hindmarsh M, Schaefer J, Bonomi A.** Improving chronic illness care: translating evidence into action. Health Aff (Millwood). 2001;20(6):64–78.

Journal of Immigrant & Refugee Studies, 8:241–260, 2010

ISSN: 1556-2948 print / 1556-2956 online
DOI: 10.1080/15562948.2010.501279

Routledge
Taylor & Francis Group

Cross-Cultural Considerations in the Recruitment of Latinos of Mexican Origin into HIV/AIDS Clinical Trials in the U.S.-Mexico Border Region: Clinician and Patient Perspectives

MARÍA LUISA ZÚÑIGA and ESTELA BLANCO

Division of International Health & Cross Cultural Medicine, University of California, San Diego, La Jolla, California, USA

LAWRENCE A. PALINKAS

School of Social Work, University of Southern California, Los Angeles, California

STEFFANIE A. STRATHDEE

University of California, San Diego, La Jolla, California, USA

ALLEN L. GIFFORD

Department of Veterans Affairs, Center for Health Quality, Outcomes & Economics Research, Bedford, Massachusetts, USA and Boston University, Boston, Massachusetts, USA

María Luisa Zúñiga, PhD, is Assistant Professor, Division of International Health & Cross Cultural Medicine, University of California, San Diego. Estela Blanco, BA, is Project Coordinator, Division of of International Health & Cross Cultural Medicine, University of California, San Diego. Lawrence A. Palinkas, PhD, is affiliated with the School of Social Work, University of Southern California. Steffanie A. Strathdee, PhD, is Professor and Harold Simon Chair Chief, Division of International Health & Cross Cultural Medicine, Department of Family and Preventive Medicine, University of California, San Diego. Allen L. Gifford, MD, is Associate Director, Department of Veterans Affairs, Center for Health Quality, Outcomes & Economics Research, and Associate Professor of Public Health and Medicine, Boston University.

Address correspondence to María Luisa Zúñiga, PhD, Assistant Professor, Division of International Health & Cross Cultural Medicine, University of California, San Diego, 9500 Gilman Dr., Dept. 0927, La Jolla, CA 92093-0927. E-mail: mzuniga@ucsd.edu

This work would not have been possible without trust and support of persons living with and/or impacted by HIV, and the clinicians and staff who dedicate their work to serving persons living with HIV/AIDS. The authors gratefully acknowledge the essential participation and contribution of Dania Barroso Conde; *Amigos de CASA*, San Ysidro Health Center, the University of California, San Diego Mother, Child, and Adolescent HIV Program, the HIV Neurobehavioral Research Center and Antiretroviral Research Center which made this study possible. This study was funded by the National Institutes of Mental Health (K01 MH072353) and also supported by the San Diego EXPORT Center, National Center of Minority Health and Health Disparities, National Institutes of Health (P60 MD00220). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.

Despite increasing prevalence of HIV among U.S. Latinos, participation in HIV clinical trials is low. Barriers to HIV clinical trial participation in U.S. Latinos are not well understood. Using in-depth, semistructured interviews with HIV care providers serving HIV-positive Latinos and focus groups with HIV-positive Mexican-origin Latinos, we assessed cross-cultural barriers (e.g., stigma and linguistic) to HIV clinical trials in San Diego, California, bordering Mexico. Cross-cultural barriers were explored using grounded theory analytical techniques. Patient-provider concordance on the nature of HIV-related stigma, linguistic barriers, the impact of U.S.-Mexico border on Latino patients and participation in clinical trials were found. Providers described care access challenges faced by patients of Mexican-origin, particularly in light of immigration and U.S. border policy. HIV-related stigma and communication barriers among Latinos remain important obstacles to clinical trials participation and care access in the United States.

KEYWORDS U.S.-Mexico border, immigrants, HIV/AIDS, health care access, clinical trials

To live on the border is to live in the center: to be at the entrance and the exit; to inhabit two worlds, two cultures and to accept both.

—José Antonio Burciaga

INTRODUCTION

HIV prevalence is increasing in the U.S.-Mexico border region (Brouwer et al., 2006) and HIV has been identified as a health priority by the U.S.-Mexico Border Health Commission and Pan American Health Organization (United States-Mexico Border Health Commission, 2003; Pan American Health Organization [PAHO], 2007). National HIV surveillance data in the United States indicate a notable increase in the proportion of Latinos among new HIV infections and AIDS diagnoses, as well as increasing rate of HIV diagnosis among Latinos—three times higher than in non-Latino Whites in 2005 (Centers for Disease Control [CDC], 2006). Latino participation in HIV clinical trials, however, remains low (Gifford et al., 2002; Getz & Faden, 2008). Broadly defined, clinical trials are biomedical (e.g., development of new medications and treatment options or improvement of medication dosing) or behavior-related (e.g., assessment of new behavioral interventions to prevent disease transmission) research studies with individuals who are assigned by the investigator to a treatment or other intervention, and their outcomes are measured (U.S. National Institutes of Health [NIH], 2009).

Clinical trials are designed to determine new ways of improving patient health, including comparison of existing treatments to determine relative effectiveness.

Participation in clinical trials may allow for access to new therapies and treatment modalities, of particular importance for minority populations most affected by the HIV epidemic who may have limited access to care. Participation in clinical trials may not guarantee access to better treatment modalities; however, it improves a patient's access to monitoring of their disease state by clinicians specialized in the field of HIV. Equitable access to clinical trials may also improve generalizability of study findings to minority populations, and may reduce health disparities among those most affected by HIV. The National Institutes of Health (NIH) has issued a directive to promote recruitment of racial and ethnic minorities into research in order to identify potential differences in effects among ethnic/racial groups (NIH, 2001). Attention to ethnic and culturally-mediated barriers to HIV care and participation in clinical trials among Latinos in the U.S.-Mexico border region is warranted given the region's rapid demographic growth and substantial daily cross-border interaction (San Diego Association of Governments [SANDAG], 2003; U.S. Census Bureau, 2007).

Barriers to clinical trials include system-level barriers that may impede access to care and clinical trials (King et al., 2007). Examples of system-level barriers include provider preconceived notions or stereotypes about willingness to participate among Latinos and African Americans; language differences (e.g., clinicians are less likely to discuss clinical trials with patients who do not speak English); and level of effort that clinical trials personnel may make to recruit members of underrepresented communities (King et al., 2007; Stone, Mauch, Steger, Janas, & Craven, 1997; Stone, Mauch, & Steger, 1998; Stone, 2005). Ethical issues related to voluntary and informed consent, participation and coercion are appropriate areas of concern when enrolling members of traditionally underserved communities, including Latinos of Mexican origin living with HIV. To this end, Institutional Review Boards and NIH require detailed researcher explanation of how study recruitment will uphold the rights of participants including training of key personnel in protection of study participants (NIH, 2000). Provisions to improve linguistically-competent and ethically sound enrollment of traditionally disenfranchised ethnic groups include projects such as Project TRES (2009) an ethnically and linguistically-tailored ethics training course for Latino/a community-level research recruitment staff. Notwithstanding, continued attention is needed on potential ethical problems, particularly language-related issues, in recruitment of underserved populations. Our preliminary work in clinical trials participation barriers among Latinas living with HIV indicated that personal factors, such as fear, shame, and stigma are the most commonly perceived barriers to clinical trials participation (Zúñiga, Blanco, Martínez, Strathdee, & Gifford, 2007). This same study found that HIV service providers,

reported system-level barriers (Zúñiga et al., 2007). HIV-related stigma has been well documented in other studies as a significant barrier to HIV care and treatment in U.S. Latino populations (Zúñiga, Brennan, Scolari, & Strathdee, 2008; Vanable, Carey, Blair, & Littlewood, 2006), yet its role in clinical trials participation for U.S. Latinos has not been well established. Furthermore, although HIV-related stigma has been implicated as a barrier to HIV vaccine clinical trials recruitment among uninfected populations (Brooks, Newman, Duan, & Ortiz, 2007) there is limited research on the role of HIV-related stigma in clinical trials recruitment of persons living with HIV. As the prevalence of HIV increases in Latino populations along the U.S.-Mexico border region, lack of access to HIV-related care, including clinical trials, may further widen observed disparities in health.

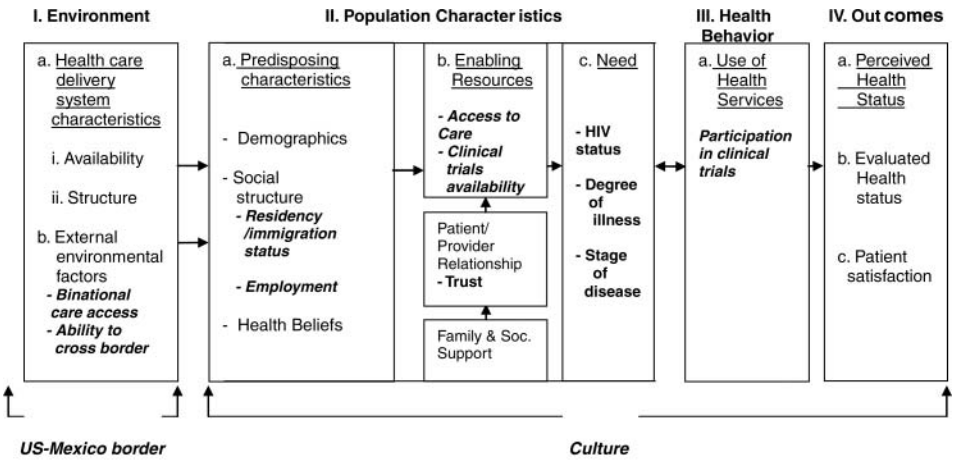
The current qualitative study was conducted in San Diego, California, which borders Tijuana, Baja California, Mexico. We examined barriers to HIV clinical trials participation among U.S. Latinos living with HIV and HIV care providers who serve Latino patients with a specific focus on stigma and system-level barriers. An in-depth understanding of patient and clinician contextual realities and perceived barriers to clinical trials is important to increase the effectiveness of clinical trials recruitment and inform the design of new approaches to improve access to HIV clinical trials and health care in Latino populations.

METHODS

Conceptual Framework

The behavioral model of health services utilization developed by Andersen, Aday, and others was adapted for the present study and served as the theoretical framework to explore and define factors associated with barriers to HIV clinical trials and HIV care in HIV-positive Latino populations in the U.S.-Mexico border region (see Figure 1; Phillips, Morrison, Andersen, & Aday, 1998; Andersen, 1995). Specifically, we applied the model's framework to include access to clinical trials as a factor of enabling resources within population characteristics (see Figure 1). The Behavioral model provides a framework in which factors associated with access to clinical trials can be contextualized within a broader access to care framework. This model has been used effectively to predict behavior in studies with HIV-positive persons and other vulnerable populations (Andersen, 1995; Gelberg, Andersen, & Leake, 2000; Dobalian et al., 2006) and was particularly well suited for the study as it has been successfully applied to access of vulnerable groups to antiretroviral therapy in the United States (Andersen, 1995).

The current study recruited two different populations (HIV providers and Latino patients living with HIV), employing sequential qualitative methods to first conduct in-depth interviews with HIV providers and later using



* Based on work from: Andersen RM, 1995; Phillips KA, Morrison KR, Andersen R, Aday LA, 1998.

FIGURE 1 Study Conceptual Framework using the Behavioral Model of Health Services Utilization.

these findings to inform focus group questions used with Latino patients living with HIV. Since there has been little research with clinicians who serve Latinos living with HIV (Zúñiga et al., 2007), it was imperative to conduct in-depth interviews with clinicians to understand their perspectives on barriers to care as part of potential structural barriers faced by patients. Furthermore, given that our previous research indicated that there may be a disconnect between perceptions of barriers to clinical trials between clinicians and patients (Zúñiga et al., 2007), we chose to conduct subsequent focus groups with Latino patients in order to allow for detailed discussion of patient agreement or disagreement with provider perspectives. This study was reviewed and approved by the University of California, San Diego (UCSD) Human Research Protection Program. Consent was obtained from all participants prior to participation.

Recruitment

In-depth, key informant interview participants included a purposive sample of 15 HIV physicians, pharmacists and nurses from affiliated university research centers, clinics and hospitals who actively served Latino populations living with HIV. Potential participants were chosen based on recommendations of key members of the UCSD HIV clinical trial system including clinical trials recruitment coordinators who work directly with clinicians. Potential key informants were contacted via e-mail and phone by a member of the research team; of the 33 persons contacted, 15 agreed to participate (45%). Our sample included clinicians who are integrally

involved with HIV clinical trials, including study principal investigators and co-investigators. Low clinician response rate is common in busy clinical environments where schedules may make it difficult for clinicians and study staff to participate; King and colleagues indicated low response rates among HIV clinical trials study staff, with rates ranging from 56% to 58% in an internet-based survey (King et al., 2007).

Focus group participants were Latino patients living with HIV in the San Diego region. Eligibility criteria included: ≥18 years of age, Latino, English- or Spanish-speaking, and living with HIV. Thirty-seven persons participated in a total of four focus groups (two male and two female groups). Of the four focus groups, two focus groups (one female and one male group) were comprised of persons who had participated in a clinical trial, and two (one female and one male group) were conducted with persons who had never participated in a clinical trial. Focus group participants were recruited by volunteer peer advocates from a partner community clinic as well as by outreach workers from the UCSD HIV clinical trial research groups.

Data Collection

The in-depth interviews were conducted using a guide that included 29 open-ended and 6 closed-ended items, based on research generated from patients living with HIV and HIV service providers that indicated specific difference in perspectives on barriers to entry into clinical trials (Zúñiga et al., 2007). Questions such as: "Do you think low acculturation influences an HIV-positive Latino's willingness or ability to participate in a clinical trial?" and "Do you think Latino patients are worried that someone will find out they are HIV-positive if they participate in a clinical trial?" wer were followed by prompts for the participants to expand on why or why not they thought the statement was correct. Trained study interviewers conducted the hour-long interviews with providers at a mutually convenient time and location. Providers did not receive an incentive for participating in the study.

Provider interview data were used to generate patient focus group questions and to ask patients about their perceptions of patient-provider communication, its role as a potential mediator of health choices, satisfaction with HIV medical services and care, and how these items may influence clinical trial participation. Draft focus group questions were presented in Spanish to a peer advocacy group for evaluation and feedback. Peer advocate review contributed to identification of errors and resulted in substantial improvements in clarity of questions in the interview measure. Focus group questions (12 total) were designed to provide participants with sufficient background information to understand the context of provider comments derived from key informant interviews and promote focus group reflection; elicit agreement or disagreement; and contribute their own thoughts on the

subject. Questions framed the issue and were followed by prompts for the participants to expand on their statements. Examples include: "For persons living with HIV who are undocumented, do you think that for those people their citizenship status is a barrier to participation in clinical trials?;" and "Do you think that the role of your family or partner could influence your participation in a clinical trial?" A 12-item demographic survey, including questions on education level, language preferences, and income, was also administered to study participants. All focus groups were conducted in Spanish and lasted between 1.5 and 2 hours; two were conducted at a collaborating community clinic and two at a UCSD clinical trial research office. Participants were read a consent form and allowed to ask questions during the entirety of the process. Participants received a $25 grocery store voucher. (Study measures and protocols are available upon request from the study's lead author).

DATA ANALYSIS

We followed the principle of *microanalysis* as described by Strauss and Corbin (1990), which is the detailed analysis of text that is necessary at the beginning of analysis to generate initial themes and suggest relationships between themes. Two members of the study team reviewed all interviews and focus group transcripts to identify themes.

Subsequent analysis was conducted independently by three study team members using open coding procedures on a sample of 5 of the 15 clinician/provider interviews to generate concepts, properties, and dimensions discovered in the data (Strauss & Corbin, 1990). The study team then met to discuss categories and reach consensus on coding themes. Next, the study team independently reviewed and coded two open-ended questions to reach consensus on how codes had been applied and whether there were emerging themes that merited new codes.

Using the revised coding scheme, two members independently coded five pages of two focus group transcripts. Coded questions were independently reviewed by third team member, who determined inter-rater reliability and met with study team to resolve coding discrepancies. Inter-rater reliability was >80%. The qualitative text analysis software, ATLAS.ti (version 7; ATLAS.ti Scientific Software Development GmbH), was used to facilitate visualization and organization of coded categories (concepts that stand for phenomena/central ideas in the data) and their properties, dimensions, and subcategories (Strauss & Corbin, 1990) as well as to facilitate generating frequencies of coded categories. Using the method of constant comparison, (Glaser & Strauss, 2006; Strauss & Corbin, 1990) a final coding taxonomy was reached through study team consensus.

RESULTS

Demographics

HIV care providers participating in in-depth interviews included 10 physicians, 3 nurses, and 2 pharmacists; were mostly non-Hispanic (66.7%), male (73.3%), and worked with HIV-positive Latinos for >10 years (73.3%). Slightly less than half of provider respondents were comfortable communicating in Spanish (47%), with one provider preferring to respond to the in-depth questions in Spanish.

Of 37 Latino focus group participants, 57% were male; and mean age was 43 years (range 22–59 years). Most participants were born in Mexico (92%) and about 60% of persons born outside of the United States had lived in the United States ≥10 years; 41% made at least one round-trip border crossing per month. About 38% had completed ≤8 years of schooling; nearly 70% reported an annual family income <$15,000/year and 27% reported no medical insurance. Most (78%) preferred receiving health information in Spanish. Eighty-eight percent of participants were exposed to HIV through sexual contact and 50% had lived with HIV for ≥10 years; about half (49%) having been diagnosed in Mexico. Additional focus group sociodemographic data are provided in Table 1.

Key Emergent Themes

Major themes identified in both in-depth provider interviews and patient focus groups were: (a) HIV-related stigma, (b) Communication/language; (c) Cross-cultural issues; and (d) U.S.-Mexico border and immigration. Findings are presented by major theme with HIV provider perspectives first, followed by patient perspectives.

HIV-Related Stigma

Providers described manifestations of HIV-related stigma as barriers to clinical trials participation among their Latino patients, noting the influence of individual-level stigma (e.g., fear of losing social support if seen entering an HIV research center) and community-level stigma (e.g., family adversely influenced their desire to seek medical care out of concern that others will learn of patient's status). Internalized stigma, both HIV-related stigma and stigma about sexual orientation, was expressed by providers who indicated that Latino patient discomfort with disclosure could serve as a barrier to participate in an HIV clinical trial. Providers conveyed that some patients felt guilty about being infected with HIV. Internalized stigma overlapped with patient concern about the number of additional clinic visits that clinical trials participation entails and having to explain to family members the reason for

TABLE 1 Demographics of Latino Focus Group Participants ($n = 37$)

Characteristics	N	%
Gender		
Male	21	56.8
Female	15	40.5
Transgender	1	2.7
Age		
18–35	7	18.9
36–51	23	62.2
52+	6	16.2
Missing	1	2.7
Family income per year		
<$15,000	26	70.3
$15,000–24,999	5	13.5
$25,000–49,999	0	0
$50,000+	1	2.7
Missing	5	13.5
Years of school completed		
0	1	2.7
1–8	13	35.1
9–12	14	37.8
>12	9	24.3
Missing	1	2.7
Country of birth		
United States	1	2.7
Mexico	34	91.9
Guatemala	1	2.7
Missing	1	2.7
Length of U.S. residency[a]		
0–5 years	7	19.4
6–9 years	6	16.7
10+ years	22	61.1
Missing	1	2.8
Monthly border crossings (round-trip)		
0	12	32.4
1–5	9	24.3
6–9	2	5.4
10+	4	10.8
Missing	10	27.0

[a]Among participants not born in the United States.

going to the doctor more often. Stigma was also raised in the context of trusting providers (e.g., patient concern over who will have access to their clinical trials information), although the broader influence of HIV stigma on the patient appeared to be more of a concern than patient mistrust of providers.

Providers felt that patients were concerned about being seen at sites where only persons living with HIV receive care. One provider indicated that a patient was anxious about being identified as someone living with HIV so the provider made recommendations as to other clinics where the patient could seek care. Co-occurring themes with HIV-related stigma were

disclosure and family social support. Although providers observed Latino patient reticence to be seen walking into an HIV research center, providers felt that it would be beneficial to have clinical trials centers closer to where Latinos reside. Some patients feared that they may be seen and somehow compromise the social support provided by their family if the patient's HIV status became known. As one provider described:

> But the problem is that they [families] are supportive [and] they try to cover the issue; one of the young ones die of HIV they will say they die of pneumonia . . . or something other than complications related to HIV. If you tell them it is an HIV clinic then they will try to skip appointments. If you say that this is a place they do HIV studies they will be labeled and for sure try to avoid us.

The increase in number of care visits related to clinical trials participation was also perceived as problematic for patients who are concerned about stigma and having to explain to family and/or employer the reason for increase in visits. Increase in number of visits associated with clinical trials participation was specifically noted as a problem for employment, where patients feared having to disclose to employers the reason for needing time off for additional medical appointments and for school, where families may feel uncomfortable about explaining to teachers why the child will need to visit the doctor so often.

When the HIV-related stigma question was posed to Latino focus group participants, two individuals in different groups asked for clarification of the word *stigma*. The facilitator allowed participants who were familiar with this term to define it for their peers in order to observe how HIV-related stigma is understood and identified within the community. For example, in the women's group one participant stated, "Stigma is, for example, when you are ashamed to talk about it, you don't want to talk about it as if it were taboo for you or others."

Upon clarification of the term, participants engaged in discussion of their personal experiences with how perceived and felt HIV-related stigma and discrimination had affected their lives and that of their families. Female participants discussed how they have coped with HIV-related stigma over time. Upon reflecting on her experiences of living with HIV, navigating feelings of stigma, and HIV clinical research, one participant noted:

> At first, HIV stigma makes one starts to become paranoid . . . you receive a letter, "we invite you to participate in a study of HIV," and you want to burn the page or rip it up due to fear of that . . . and then you start to close your doors, when it should not be like that. You should talk with your family and overcome that barrier, but I'm saying not with all the family, but with close family, we should do that.

Women in both focus groups discussed at great length their efforts to avoid HIV-related stigma, mentioning having to be dishonest with family or persons in their social and faith networks to avoid being stigmatized by family and their community. Latinas discussed naming other diseases they or their partners had to avoid HIV disclosure (e.g., cancer). Important to the discussion was past experiences with discrimination related to living with HIV and being singled out with questions about how they were infected; for married women this was a particularly difficult point, navigating both their partner's HIV status as well as their own. Men did not specifically mention masking their HIV infection by telling friends and family they had another disease. One male participant stated, "Almost all of my neighbors know I have AIDS; to me, it doesn't matter, it doesn't matter to me what people say." The same participant however did admit that the stigma related to having HIV was a problem and concern for many other people: "It depends on the person . . . but yes there are a lot of people it [stigma] affects . . ."

Focus group participants also mentioned they often avoided discussion of HIV because they were tired of sharing their experiences with others and having to field more questions. However, apart from one participant's general concern about maintaining the confidential nature of HIV-related medical appointments, focus group participants did not otherwise mention concern about being seen at an HIV medical clinic and were enthusiastically in favor of having clinical trials sites closer to where they live.

Communication/language

Provider discussion of communication and language issues centered on the desire for improved access to bilingual staff and interpreters to facilitate enrollment of Latino patients with limited English language proficiency into clinical trials. Some providers expressed frustration with having a Spanish-speaking Latino patient, who was eligible for a study, but lacked access to an interpreter and having to ask the patient to wait. Some providers who did not speak Spanish were either learning or expressed interest in opportunities to improve Spanish-language skills.

Focus group participants talked about challenges to effective communication with health care providers as well as experiences that optimized communication. Although the importance of learning English was generally acknowledged, both men and women's groups expressed their preference communicating in Spanish during their medical appointments. Participants noted as well that although they may grasp working knowledge of English, their exposure to English medical terminology may be limited. In both male and female groups, language was also raised in the context of having trust in one's provider, and that a good relationship with the provider included ability to communicate in Spanish. Language was viewed by some as a

mediator of satisfactory clinical encounters, including availability of interpreter and patients expressed appreciation of clinician efforts to learn Spanish. As one patient expressed: "My doctor is learning Spanish and I love that . . ."

For some participants, limited English proficiency (speaking and reading) was a barrier to clinical trials participation, and one female participant stated this remained a barrier even when she offered to bring someone who could interpret during appointments, she stated:

> For example, I wanted to enroll into a neuropathy study that interests me a great deal because I have neuropathy, and they told me: "you know what, um, but you need to be proficient in English." I say: "what if I bring an interpreter?. . ." [provider response]: "No, you have to speak it."

Language was not reported by men as a specific barrier to clinical trials, but the importance of having access to interpreters was very important to both male and female focus group participants. One male focus group participant described having had a highly satisfactory interaction with a clinician through an interpreter on a telephone; the interaction included a conversation about sensitive topics whereby the patient felt that a telephone interpreter was less obtrusive to effective communication than having an interpreter physically in the consultation.

Patients felt that communication with providers was compromised when providers stereotyped them and felt that their relationship with the provider was undermined when clinicians made assumptions that the patient did not feel were accurate. The issue of cross-cultural communication also emerged with reference to diet. One male focus group participant described the following comments made by his physician: "Because you Mexicans are accustomed to eating a lot of fats . . . it's that your metabolism [of Mexicans] . . . since young children you eat many fats, then it is not like the Americans who eat a little less fat." A similar sentiment was expressed by a female participant who felt that she was unable to communicate effectively with her provider that she was eating healthfully although she was experiencing weight gain.

Cross-Cultural Issues

Providers mentioned the health belief of *fatalismo* [fatalism], or the idea that all things happen by fate or are destined to happen in relation to participation in clinical trials. According to one of the providers:

> There is an element of fatalism sometimes in the Latino community about HIV. That now I have it, I'm going to die soon, and if they want me in a study that means that I really am going to die soon.

Another cultural consideration providers noted about their Latino patients is a deep respect some have for the physician. One provider reflected: "[There is] more of a dependence on provider recommendation . . . [Latinos] see [the] doctor as an authority figure . . . may feel some pressure to comply and participate." Another provider mentioned: "[Latinos] are more likely to do what the doctor recommends . . . or want to participate [in clinical trials] to keep their doctor happy." Male and female focus group participant reflections on culture included awareness of differences in communication styles and stereotyping of non-Latinos, as one Latino focus group participant stated:

> We Latinos are accustomed to having people be a little warmer . . . "how are you? . . . how has it been going?" . . . even if it is a quick chat of two minutes. . . . the culture, the Germans, the Saxons, it is true, the culture is naturally colder, not because they are bad people or anything like that.

A second male focus group participant indicated a positive experience with care he received in the United States from physicians from Tijuana: "They understand us. They know our culture, our problems." Discussion ensued on the patient's responsibility to learn about the provider's culture and one male participant recommended that U.S. providers learn about the health care delivery system in Mexico to better understand how Mexican immigrants in the United States experience care delivery in their country of origin.

U.S.-Mexico Border Issues

Providers were aware of their Latino patients' strong ties to Mexico, including residence or having family in Tijuana. Awareness included knowledge of long wait times to cross the border and patient ability to make appointments on time. Border and immigration-related concerns and patient experiences also included co-occurrence with HIV stigma:

> The ones that live in Mexico. . . . the rest of the medical system doesn't know of their diagnosis, they live in a constant state of fear that they might get sick and they might need emergency treatment, there they would be thrown in a hospital where they might have relatives, nurses or people that know them around them, they don't want to trust them over there with their diagnosis, and unfortunately that creates an issue because physicians and nurses there deal with patients that are HIV-positive without knowing that they're treating HIV. And, of course, that skews their intervention and biases their diagnostic ability.

Male and female focus group participants discussed many instances of how the U.S.-Mexico border and immigration policies impact their daily lives. Latino men discussed differential access to HIV care and medication

depending on one's ability to cross the border, and efforts to donate unused medication to HIV-positive persons living in Tijuana. Women also discussed this circumstance, adding that there was an anxiety for them when they were crossing back from Tijuana to San Diego. Women expressed fear that U.S. border patrol agents may ask what their personal medications are for, since current U.S. policy indicates that non-U.S. citizens living with HIV may be denied entry into the United States or denied U.S. visas (National Immigration Project, 2004). Women also discussed how reliance on public transportation and concern over border patrol "sweeps" on public transportation would influence their decision to participate in a clinical trial. This consideration of transportation as a barrier to care is unique to U.S. immigrants who hold certain types of visas or who are undocumented in the United States. Women also expressed keen awareness of the implications of deportation on their ability to access HIV care, indicating concern about a lack of knowledge about the types and location of HIV medical services in Tijuana. Deportation was also mentioned as a concern among women because it separates families and potential sources of support for persons living with HIV.

DISCUSSION

Our study has several important implications for Latino participation in clinical trials and access to HIV care that provide a deeper understanding of how HIV-related stigma, communication and language, and other related cross-cultural issues and the border region impact the lives of Mexican-origin immigrants living with HIV. With some important modifications to fit our border context, the behavioral model of health services utilization adapted for this study served well as an explanatory framework for participation in clinical trials (Figure 1). Model components such as health behavior (participation in clinical trials) and outcomes (patient satisfaction with the clinical encounter) indeed appear to be influenced by the environment (U.S.-Mexico border and access to HIV care in binational context); population characteristics/predisposing characteristics (e.g., patient language, immigration status) and enabling resources (e.g., patient-provider relationship, including trust and patient ability to understand what is being said during their appointments). Our findings indicate that participation in clinical trials in this population may be influenced by cross-cultural issues that span all components of the model. We also note that resolution of environmental barriers such as moving clinical trials centers closer to communities may not necessarily improve participation if patients or their families are concerned about HIV-related stigma associated with being seen at a health care facility for HIV, including clinical trials center.

Across the four focus groups there was general agreement with provider observations and in many cases, participants expanded in great length on

these observations, engaging in a lively dialogue and expressing their experiences and additional views on HIV-related coping strategies, including failed coping strategies. Patient discussions also offer insights on stereotyping; feeling stereotyped and how patients may stereotype non-Latinos, both of which enrich understandings of how patients may respond to different clinical encounters. Focus group discussions also revealed patient sensitivity to issues surrounding fear of deportation, raised specifically during discussion of transportation-related barriers to care. This finding may serve to improve provider communication with their Latino patients who rely on public transportation.

Our findings extend those of previous studies that have quantified barriers such as lack of information about clinical trials; lack of transportation; limited clinic hours; concern over being part of an experiment; and fear of adverse events from treatment (Stone et al., 1997; Zúñiga et al., 2008; Travieso, 2003). In the current study, HIV care providers articulated several instances of how HIV-related stigma directly impacted Latino patient participation in clinical trials, particularly when patients' families are either not aware of the person's HIV status, or are aware but want to protect the patient from HIV-related stigma and may discourage patient participation in a clinical trial. The role of family appears to have a potential duality in the context of clinical trials, where families who serve as a source of social support and protectors of persons living with HIV may in some cases not support participation when HIV-related stigma is a concern. The nature of HIV stigma and how it permeates the lives of Latino patient study participants provided a deeper understanding of the patient context in which care decisions, including clinical trials participation, are made. Our work indicates that HIV medical care providers are attuned to the impact of stigma on the lives of their patients, but further study is needed to raise awareness of how HIV-related stigma may play a direct role in clinical trials participation. Gender-based differences in the impact of HIV-related stigma and coping mechanisms are also important to explore. For example, women in our focus groups mentioned coping with HIV-related stigma by telling friends and/or family members that they had cancer, rather than HIV. Male focus group participants did not identify this as a coping strategy.

Both patients and providers underscored the impact of HIV-related stigma in immigrant patients who currently live in or frequently visit Mexico. Concerns about patient lack of disclosure to other health care providers in Mexico or fear of their HIV status being disclosed if care were received in Tijuana were raised by providers and not patients. However, wherever care is received, both patients and providers mentioned concern about being recognized at an HIV care facility. In contrast to other studies that identified provider language-mediated barriers to clinical trials participation (King et al., 2007; Stone et al., 1998), our study indicated a genuine earnestness on the part of some providers to accommodate for patient language needs.

Providers were aware of the barriers that limited English proficiency could place on patient ability to participate in clinical trials and provided recommendations on how to reduce these barriers. Patient resourcefulness to bring their own interpreter, however, was not necessarily a bridge for this barrier when clinical trials materials were available in English only.

Our finding on the role of patient-provider racial or ethnic concordance offers partial support to the observation of Sohler and colleagues (Sohler, Fitzpatrick, Lindsay, Anastos, & Cunningham, 2007), that patient-provider ethnic or racial concordance was not associated with trust in provider, that is, patients tended to trust their providers regardless of whether they were of the same race/ethnicity or not. Patients in our study were aware that culturally-effective clinical communication is enhanced when receiving care from clinicians who are of Mexican-origin and familiar with Latino culture, however they were particularly appreciative of efforts made by their non-Latino providers to learn Spanish.

We observed several instances of the seemingly paradoxical nature of improving inclusivity of Latino populations into HIV clinical research studies and HIV care provision in a border context. Paradoxical responses were found for clinical trials locations; the role of family in supporting clinical trials participation; and delivery of culturally-effective care (Zúñiga et al., 2006). Although providers and patients indicated enthusiastic endorsement of having clinical trials recruitment and care centers closer to where Latino communities reside, stigma, manifested as concern over being seen at a clinical trials site or clinic exclusively for persons living with HIV, was expressed during interviews with providers as a clear barrier to clinical trials participation.

Prior recommendations to reduce barriers to HIV clinical trials participation for traditionally underserved communities include improving patient satisfaction with the quality of the patient-provider relationship and improving physician referrals to clinical trials (Cargill & Stone, 2005). Based on our study findings, opportunities to move HIV clinical trial centers closer to communities may address logistical problems for many Latinos living with HIV. However, this will necessitate thoughtful consideration of how stigma can be mitigated, with the essential need for feedback from community members themselves. Community member perspectives on research, the research institution's reputation in the community, as well as the relationship between researchers and the target community are all important considerations for successful recruitment into research studies (Sullivan et al., 2001). Concern over immigration issues necessitates that clinicians who serve immigrant communities also consider this as a potential barrier to care and clinical trials access for their patients. Provider opportunities to mitigate this concern may include becoming informed about immigrant patient rights (National Immigration Project, 2004); being sensitive to the potential for

immigration-specific duress in immigrant populations (e.g., concern about taking public transportation to clinical trials site for fear of encountering immigration authorities); and reassuring patients that their relationship with providers will not be compromised through a violation of trust.

Because our populations were drawn from care centers affiliated with a university research center, our findings may not be generalizable to other providers and Latino subgroups or nonborder dwelling populations, or populations who have had less experience with clinical trials. Although study researchers made an effort to recruit persons living with HIV who had and had not participated in clinical trials, it was apparent, particularly in the men's groups, that most participants had some experience with and knowledge about clinical trials due to receiving care from providers affiliated with the academic medical center and perhaps also due to length of time living with HIV. Nevertheless, future study is needed to understand the relative importance of major barriers in Latino populations. We also wish to acknowledge potential bias in focus group study questions. For example, questions specific to documentation status as a barrier to care could have elicited biased responses. We found, however, that as a whole, our study measure elicited frank participant discussion of sensitive issues such as stress of crossing the border with HIV medications or fear of deportation which enriched the depth of understanding of their perceived barriers to clinical trials participation and care. Finally, our small sample size may have missed important issues that larger studies may have uncovered with patients and HIV clinicians. The qualitative nature of this work, however, provides us with an in-depth understanding of barriers to clinical trials that informs future interventions to improve clinical trials representation.

CONCLUSIONS

Understanding barriers to participation in clinical trials is fundamentally inseparable from barriers to HIV-related care faced by Latino populations overall. Efforts to decrease disparities in clinical trials participation among Latino populations will include many, if not all, of the same strategies to improve patient engagement into care. This study attempts to fill an important gap in our understanding of low Latino recruitment into clinical trials, and is worthy of attention in future patient-level or provider-level interventions to promote clinical trials participation. Efforts must be made to improve consciousness among HIV clinical trials clinicians and staff of the multiple barriers potentially faced by their Latino patients, including the profound impact of HIV stigma and limited English proficiency on clinical interactions and activities. That immigrants of Mexican origin comprise a large proportion of persons living with HIV/AIDS in California makes imperative the inclusion of

deportation anxiety and U.S. immigration policies in discussions of improving access to HIV care, including participation in clinical trials.

REFERENCES

Andersen, R. M. (1995). Revisiting the behavioral model and access to medical care: Does it matter? *Journal of Health and Social Behavior*, *36*, 1–19.

Brooks, R. A., Newman, P. A., Duan, N., & Ortiz, D. J. (2007). HIV vaccine trial preparedness among Spanish-speaking Latinos in the US. *AIDS Care*, *19*, 52–58.

Brouwer, K. C., Strathdee, S. A., Magis-Rodríguez, C., Bravo-García, E., Gayet, C., Patterson, T. L., (2006). Estimated numbers of men and women infected with HIV/AIDS in Tijuana, Mexico. *Journal of Urban Health: Bulletin of the New York Academy of Medicine*, *83*, 299–307.

Cargill, V. A., & Stone, V. E. (2005). HIV/AIDS: A minority health issue. *The Medical Clinics North America*, *89*, 895–912.

Centers for Disease Control. (2006). HIV/AIDS among Hispanics—United States, 2001–2005. *MMWR*, *56*, 1052–1757. Retrieved from http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5640a4.htm

Dobalian, A., Andersen, R. M., Stein, J. A., Hays, R. D., Cunningham, W. E., & Marcus, M. (2003). The impact of HIV on oral health and subsequent use of dental services. *Journal of Public Health Dentistry*, *63*(2), 78–85.

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 Services Research*, *34*, 1303–1305.

Getz, K., & Faden, L. (2008). Racial disparities among clinical research investigators. *American Journal of Therapeutics*, *15*(1), 3–11.

Gifford, A. L., Cunningham, W. E., Heslin, K. C., Andersen, R. M., Nakazono, T., Lieu, D. K., (2002). Participation in research and access to experimental treatments by HIV-infected patients. *New England Journal of Medicine*, *346*, 1373–1382.

Glaser, B. G., & Strauss, A. L. (2006). *The discovery of grounded theory: Strategies for qualitative research*. New Brunswick, NJ: Aldine Transaction.

King, W. D., Defreitas, D., Smith, K., Andersen, J., Patton, P. L., Adeyemi, T., (2007). Attitudes and perceptions of AIDS clinical trials group site coordinators on HIV clinical trial recruitment and retention: A descriptive study. *AIDS Patient Care and STDs*, *21*, 551–563.

National Immigration Project. (2004). *HIV/AIDS and immigrants: A manual for HIV/AIDS service providers*. Boston, MA: Author.

National Institutes of Health. (2000). *Required education in the protection of human research participants*. Retrieved from http://grants.nih.gov/grants/guide/notice-files/NOT-OD-00-039.html

National Institutes of Health. (2001). *NIH policy and guidelines on the inclusion of women and minorities as subjects in clinical research—amended 2001*. Retrieved from http://grants.nih.gov/grants/funding/women_min/guidelines_amended_10_2001.htm

Pan American Health Organization. (2007). Health in the Americas, 2007. Volume II—Countries. United States-Mexico Border Area. Retrieved from http://www.crid.or.cr/digitalizacion/pdf/eng/doc16714/doc16714-13.pdf

Phillips, K. A., Morrison, K. R., Andersen, R., & Aday, L. A. (1998). Understanding the context of healthcare utilization: Assessing environmental and provider-related variables in the behavioral model of utilization. *Health Services Research*, *33*, 571–596.

Project TRES. (2009). Training in research ethics and standards. Retrieved from http://www-rohan.sdsu.edu/~gra/grad/research/projecttresinfo.html

San Diego Association of Governments (SANDAG). (2003, July). *San Diego-Baja California land ports of entry fact sheet*. San Diego, CA: Author.

Sohler, N. L., Fitzpatrick, L. K., Lindsay, R. G., Anastos, K., & Cunningham, C. O. (2007). Does patient-provider racial/ethnic concordance influence ratings of trust in people with HIV infection? *AIDS Behavior*, *11*, 884–896.

Stone, V. E. (2005). Physician contributions to disparities in HIV/AIDS care: The role of provider perceptions regarding adherence. *Current HIV/AIDS Reports*, *2*, 189–193.

Stone, V. E., Mauch, M.Y., & Steger, K. A. (1998). Provider attitudes regarding participation of women and persons of color in AIDS clinical trials. *Journal of Acquired Immune Deficiency Syndrome & Human Retrovirology*, *19*, 245–253.

Stone, V. E., Mauch, M. Y., Steger, K., Janas, S. F., & Craven, D. E. (1997). Race, gender, drug use, and participation in AIDS clinical trials: Lessons from a municipal hospital cohort. *Journal of General Internal Medicine*, *12*, 150–157.

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Thousand Oaks, CA: Sage.

Sullivan, M., Kone, A., Senturia, K. D., Chrisman, N. J., Ciske, S. J., & Krieger, J. W. (2001). Researcher and researched-community perspectives: toward bridging the gap. *Health Education & Behavior*, *28*, 130–149.

Travieso, P. D. (2003). Qualitative research through focus groups on Ryan White title IV consumer fears and perceived barriers to clinical research. Cooperative agreement to provide training and technical assistance to title IV grantees (Unpublished data). University of Miami Department of Family Medicine and Community Health.

United States-Mexico Border Health Commission. (2003). *Healthy border 2010: An agenda for improving health on the United States-Mexico Border*. Retrieved from http://www.borderhealth.org/files/res_63.pdf

U.S. Census Bureau. (2007). *Press release: Minority population tops 100 million*. Retrieved from http://www.census.gov/newsroom/releases/archives/population/cb07-70.html

U.S. National Institutes of Health: ClinicalTrials.gov. (2009). *Understanding clinical trials*. Retrieved from http://clinicaltrials.gov/ct2/info/understand

Vanable, P. A., Carey, M. P., Blair, D. C., & Littlewood, R. A. (2006). Impact of HIV-related stigma on health behaviors and psychological adjustment among HIV-positive men and women. *AIDS and Behavior*, *10*, 473–482.

Zúñiga, M. L., Blanco, E., Martínez, P., Strathdee, S. A., & Gifford, A. L. (2007). Perceptions of barriers and facilitators to clinical trials participation in HIV-positive Latinas. *Journal of Women's Health*, *16*, 1322–1330.

Zúñiga, M. L., Brennan, J., Scolari, R., & Strathdee, S. A. (2008). Barriers to HIV care in the context of cross-border health care utilization among HIV-positive persons living in the California/Baja California US-Mexico border region. *Journal of Immigrant and Minority Health*, *10*, 219–227.

Zúñiga, M. L., Sidelinger, D. E., Blaschke, G. S., Silva, F. A., Broyles, S. L., Nader, P. R., (2006). Evaluation of residency training in the delivery of culturally effective care. *Medical Education*, *40*, 1192–1200.